LATENCY AS A FUNCTION OF NUMBER OF RESPONSE ALTERNATIVES IN PAIRED-ASSOCIATE LEARNING

W. K. Estes

and

Donald P. Horst

TECHNICAL REPORT NO. 135

July 1, 1968

PSYCHOLOGY SERIES

This research was supported in part by Grant GB-3878 from the National Science Foundation and by Grant MH-6154 from the United States Public Health Service. The authors wish to acknowledge the contributions of Professor R. C. Atkinson to the planning of the experiments reported.

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Abstract

Two experiments were planned to elucidate, firstly, the extent to which changes in response latency during paired-associate acquisition reflect an associative learning process and, secondly, whether latencies at any given stage of learning vary with number of items and, for a constant number of items, with number of response alternatives, in the manner expected if response selection is primarily a search process.

In Experiment 1, subjects learned two successive paired-associate lists with the same response sets, each list comprising three sublists of 8 stimuli with 2, 4, or 8 response alternatives. The principal finding was that learning curves for both correct and error latencies were virtually identical for the first and second lists; the large and systematic decreases in latency over trials in each instance could not be attributed to such variables as response availability and must be taken to reflect the associative learning process. Further, the pattern of learning curves for the three sublist conditions was similar in all essentials for error frequency curves and response latency curves.

In Experiment 2 subjects were given extended training on a 6-and a 24-item list, each list having a different correct response for each stimulus. Both correct response and error latency curves declined to common asymptotes which differed by less than .05 sec. for the two lists.

In neither experiment did terminal response latencies vary with number of items or with number of response alternatives to a degree which could be taken to support the hypothesis that asymptotic paired-associate latencies reflect a search process.

A stimulus sampling model accounts semiquantitatively for the main findings of both experiments, but mathematical problems involved in computing detailed numerical predictions remain to be solved.

Latency as a Function of Number of Response Alternatives in Paired-Associate Learning

W. K. Estes

and

Donald P. Horst

Stanford University

It is commonly assumed in behavior theory (e.g., Estes, 1959a; Hull, 1943; Spence, 1951) that measures of response probability and latency are alternative, and essentially equivalent, indices of learning. However, the assumption has been adequately documented by empirical evidence only for certain elementary animal learning situations. Recently the same assumption has begun to appear as a basis for predictions about human verbal learning, though without explicit attention to the kinds of evidence needed to justify the extension. For example, Eimas and Zeaman (1963), Kintsch (1965), Millward (1964), and Williams (1962) have used latency data as a source of evidence concerning all-or-none vs. incremental interpretations of paired-associate learning. Crothers, Suppes, and Weir (1962) used response latency as a measure of item difficulty in a Russian vocabulary learning experiment. Suppes, Groen, and Schlag-Rey (1965) incorporated assumptions about latencies in a three-state Markov model for paired-associate learning; they assumed that a different response latency distribution is associated with each discrete learning state and that when the subject in a given state is presented with a stimulus member of an item his latency is drawn randomly from the appropriate distribution.

In view of the evidently increasing importance of response times as a tool for the analysis of more complex forms of learning, it seems desirable to determine the extent to which changes in latency during learning reflect associative processes as opposed to those customarily subsumed under such concepts as warm-up, changes in response availability (Horowitz 1966), or adaptation to the experimental situation. The first experiment to be reported was addressed especially to this point. The subjects were given training on two successive paired-associate lists in the same situation, with the same set of responses and in part the same stimuli being carried over from the first list to the second though with new stimulus-response assignments.

Decreases in latency were expected to occur throughout learning of the first list, whether as a result of general adaptational factors and the like or as a reflection of associative learning. The curve for latencies as a function of trials on the second list was expected to provide evidence differentiating these possible sources of variation. If the initial level of latencies on the second list should be substantially lower than at the beginning of the first list, we would infer that part at least of the decrease in latencies during first list learning must be a function of such factors as warm-up and changes in response availability which would not recur during second list learning in the same situation. If the initial level and the rate of decline of latencies during second list learning should be essentially the same as during first list learning, we would infer that the changes in latency are primarily a reflection of associative learning processes.

Our second main interest in response times concerns their possible usefulness as a tool for analyzing retrieval processes in recall. If recall involves some type of search process, in which the set of possible responses to a presented stimulus is brought into memory and scanned until the correct one is identified, latency of recall should be directly related to the number of alternative responses associated with the stimulus. The lists used in this experiment were constructed so as to permit determinations of this relationship, with controls of some variables which vary concomitantly with the number of response alternatives in standard experiments dealing with measures of learning as a function of list length.

Method

Design. Each subject learned two 24-item lists of paired-associate items having letter pairs as stimuli and digits as responses. The subjects were run 12 trials per day for 5 days on List A, then 12 trials per day for 5 days on List B. Each list was made up of 3 sublists as indicated in the following example:

List A			List B		
Sublist 8	Sublist 4	Sublist 2	Sublist 8	Sublist 4	Sublist 2
GH - 8	TH - 8	XH - 1	GV - 4	TV - 8	XV - 1
GQ - 1	TQ - 6	XQ - 6	GK - 5	TK - 3	XK - 6
GD - 1	TD - 3	XD - 1	GN - 8	TN - 6	XN - 6
GZ - 3	TZ - 1	XZ - 6	GF - 1	TF - 1	XF - 1
GB - 6	TB - 1	XB - 1	GJ - 6	TJ - 6	XJ - 6
GW - 5	TW - 8	XW - 6	GL - 3	TL - 1	XL - 1
GM - 7	TM - 6	XM - 6	GR - 2	TR - 3	XR - 1
GC - 4	TC - 3	XC - 1	GS - 7	TS - 8	XS - 6

Sublist 8 had eight response alternatives. Sublist 4 had four alternatives, two of which were used in Sublist 2. The first letter of the stimulus indicated the sublist to which the item belonged. The same first letters were used in List A and List B and were associated with the same response sets in both. Two non-overlapping sets of eight letters each were used for the second letters in the stimuli, one set for each list. In other words, List A and List B differed only in the second letters of the stimuli.

A unique stimulus set was prepared for each of the 12 subjects. For each subject a random permutation of the twenty consonants was generated. The first three letters in each permutation were used as the "first letters" for the 3 sublists (G, T and X in the example above). The next 8 letters were used as "second letters" in Phase 1 (H, Q, D, Z, B, W, M, C in the example), and the next 8 letters as the second letters in Phase 2 (V, K, N, F, I, L, R, S in the example). The twentieth letter in each permutation was not used.

Four response assignments were used with 3 subjects per assignment. These assignments are as follows

	Sublist 8	Sublist 4	Sublist 2
Assignment 1	1, 2, 3... 8	1, 3, 6, 8	1, 6
Assignment 2	1, 2, 3... 8	1, 3, 6, 8	3, 8
Assignment 3	1, 2, 3... 8	2, 4, 5, 7	2, 5
Assignment 4	1, 2, 3... 8	2, 4, 5, 7	4, 7

The sample S-R list given above illustrates one application of Assignment 1.

The 3 sublists in each list were combined to form a single 24 item list. Each trial consisted of a random permutation of the 24-item list. There was no break between trials, and the subjects were not given any information about the structure of the lists.

It was assumed that during training on List A, subjects would learn the response sets associated with the initial stimulus letters. Then, in terms of the example above, upon presentation of a stimulus with initial letter T, the subject would have to select his response from the subset 1, 3, 6, 8; and upon presentation of a stimulus with an initial X, he would have only to choose between 1 and 6. But, since the subject could not know prior to any stimulus presentation what condition would be represented, the comparison of sublists with different numbers of response alternatives would not be confounded with differences in such factors as preparatory set.

Subjects. Twelve Stanford University students were used as subjects. They were each paid $20 at the conclusion of the experiment.

Apparatus. The experiment was conducted in a $13\frac{1}{2}$ ft. by 10. ft. room containing two 2 ft. by 3 ft. tables $30\frac{1}{2}$ in. high, on which were placed the stimulus-reinforcement presentation units, and the response panel. The stimulus presentation unit was a box 14 in. wide, 10 in. deep and 4 in. high. On the front of this box was a 2 in. by 12 in. screen divided vertically into eight 2 in. by $1\frac{1}{2}$ in. character positions. Each character position contained a pattern of 14 filaments. By lighting appropriate combinations of these filaments it was possible to display representations of alphabetic, numeric and special characters, although there was some overlap between the alphabetic and numeric character sets (for example, the letter "S" and the number "5" were the same). The stimulus unit was placed on top of the reinforcement unit. The reinforcement unit was identical to the stimulus presentation unit and was used to display the correct responses.

On the table at which the subject was seated was the response panel, an aluminum box 12 in. by 17 in. and 3 in. high. On the top of the box there were two parellel rows of buttons 2 in. apart. The near row was numbered from 1 to 8. The far row was not used in Experiment 1 and was covered by a strip of cardboard. Each button was $\frac{7}{8}$ in. wide and 1 in. high. There was a $\frac{1}{8}$ in. strip between adjacent buttons and the buttons projected $\frac{1}{8}$ in. above the surface of the panel. The numbers were $\frac{1}{2}$ in. block numerals on the surface of the box immediately beyond the near row of buttons. Two inches below button number six was a red plastic button $\frac{1}{2}$ in. square and $\frac{3}{4}$ in. high. This button was not used for Experiment 1.

Procedure. Subjects rested their fingers on the nearer row of buttons much as if they were typing. Thumbs were not used. When seated at the response panel the distance from the subjects' eyes to the screens was approximately 65 in.

For each subject the stimulus-response list was punched on an IBM keypunch, one S-R pair per card. An IBM 7090 computer was used to translate the characters on these cards into the multiple punched format required by the stimulus presentation system and to punch out one random permutation of the list for each trial. The stimulus-response cards were read by a modified IBM 526 keypunch into the storage units of a programming unit, which activated the stimulus and response screens at the appropriate times. The programming unit, built by Stanford Research Institute, incorporated banks of relays as storage units, electronic timers to control the rate of presentation, and a Beckman Universal EPUT and Timer Model 5230RP to measure latencies. The Beckman timer was adjusted so that all latencies greater than 9.99 sec. were recorded as 9.99 sec.

The flow of one stimulus presentation cycle proceeded as follows: A stimulus-response pair was read by the IBM 526 keypunch from columns 3-34 of a stimulus card. The card stopped at column 35 and the stimulus was presented immediately. The onset of the stimulus started the Beckman timer on the programming unit. The subject responded by pressing one button. The button lighted up when it was pressed. When the button was pushed the timer was stopped and the correct response was presented on the lower screen. The stimulus and response remained on together for 2 sec. Then the light in the button went out and both screens went blank for 3 sec. before the next stimulus was presented. During the blank period the response and the latency (truncated to hundredths of a second) were punched into the card at the punch station of the keypunch in columns 37-41. This card then passed to the read station where the S-R pair was read from columns 3-34 to start the next trial.

The keypunch and programming unit were in a room adjoining the experimental room. A two-way, sound insulated window connected the rooms and allowed the subject to indicate to the experimenter when he was ready to begin.

The essentials of the instructions to the subjects were as follows: "You are to learn to associate numbers with pairs of consonants. A pair of consonants will be shown on the upper screen. You are to make your best guess as to the number that goes with the letters and press the corresponding button. As soon as you press any button, it will light up and the correct number will appear on the lower screen. The number on the lower screen is always the correct number no matter which button you have pressed and you must press some button before the correct number will appear. Rest your eight fingers on the eight-numbered buttons at all times. You will be given a short break in the middle of the experiment during which both screens will be blank and you can move around. We are interested in both the accuracy and the speed of your responses. Though you will be only guessing at first, as you learn the correct numbers please respond as rapidly as you can without making mistakes." On successive days subjects were reminded to be as fast and accurate as possible. On Day 6 they were told that they would have to learn a new list but no information was given concerning similarity between the new and old list. On days within a phase subjects were told that the list would be the same as that given on the preceding day.

Results

Mean learning curves by daily blocks for each sublist are presented for error proportions in Figure 1, for correct response latencies in Figure 2, and for error latencies in Figure 3. Comparison of the corresponding curves for List A and List B in the upper and lower panels, respectively, of Figure 1 appears on the surface to indicate a substantial

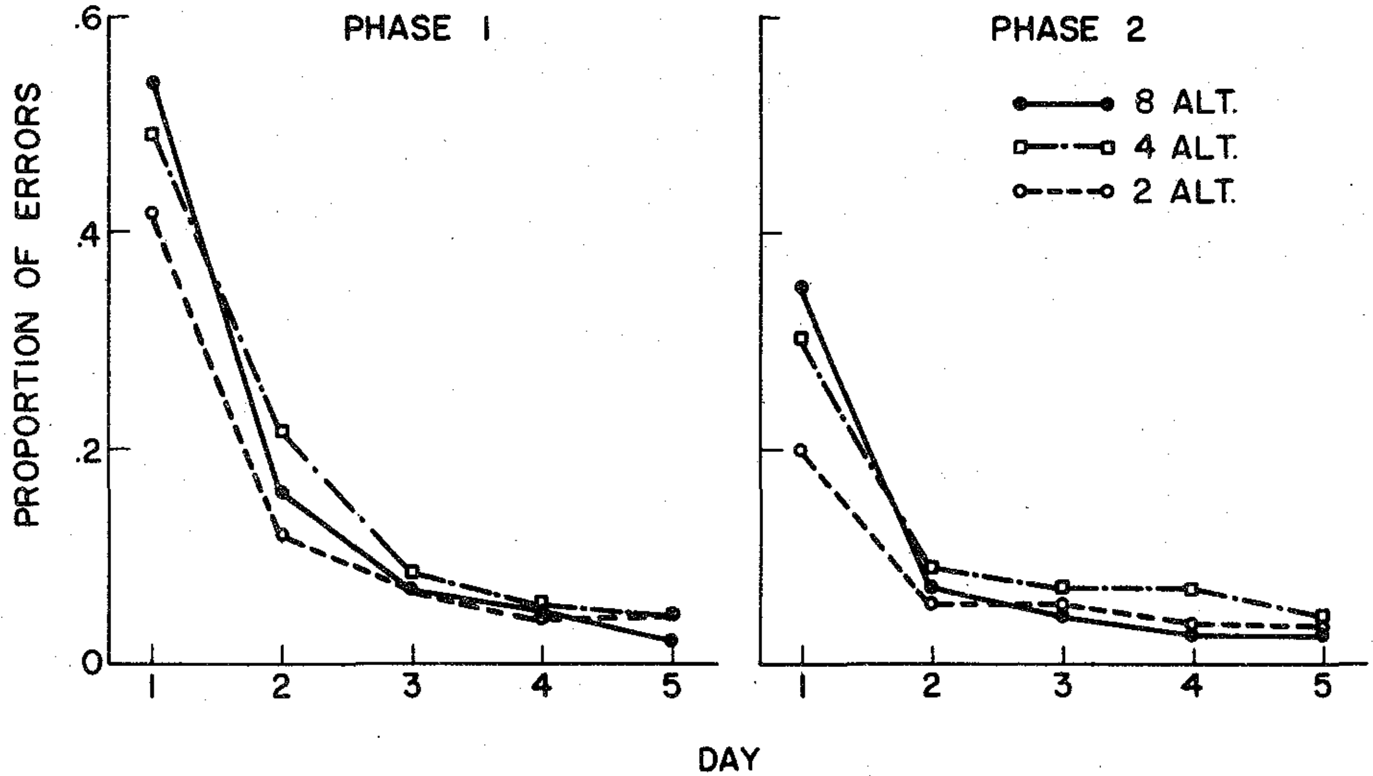
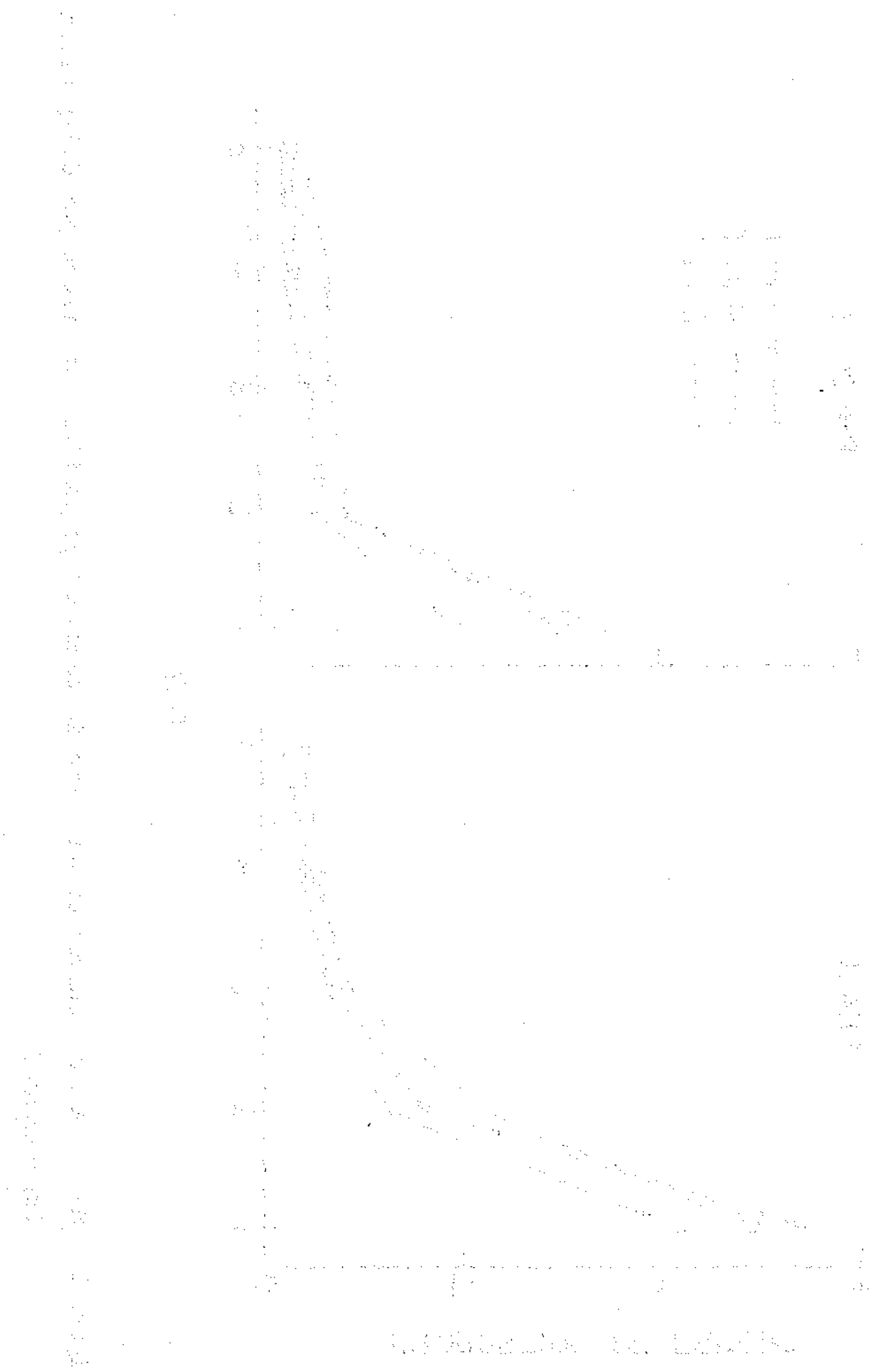


Figure 1. Mean learning curves by sublists in terms of proportions of errors per daily block, Experiment 1.



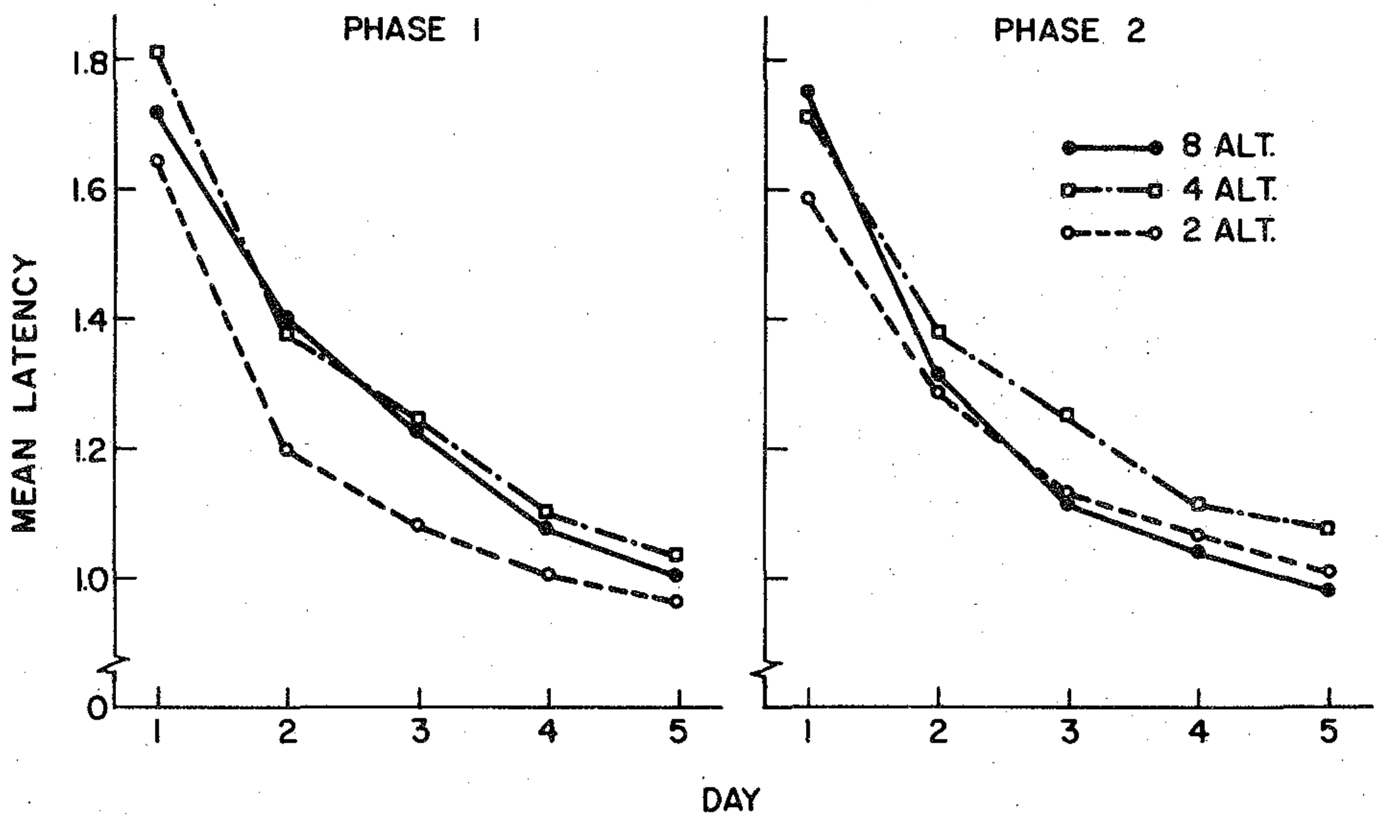
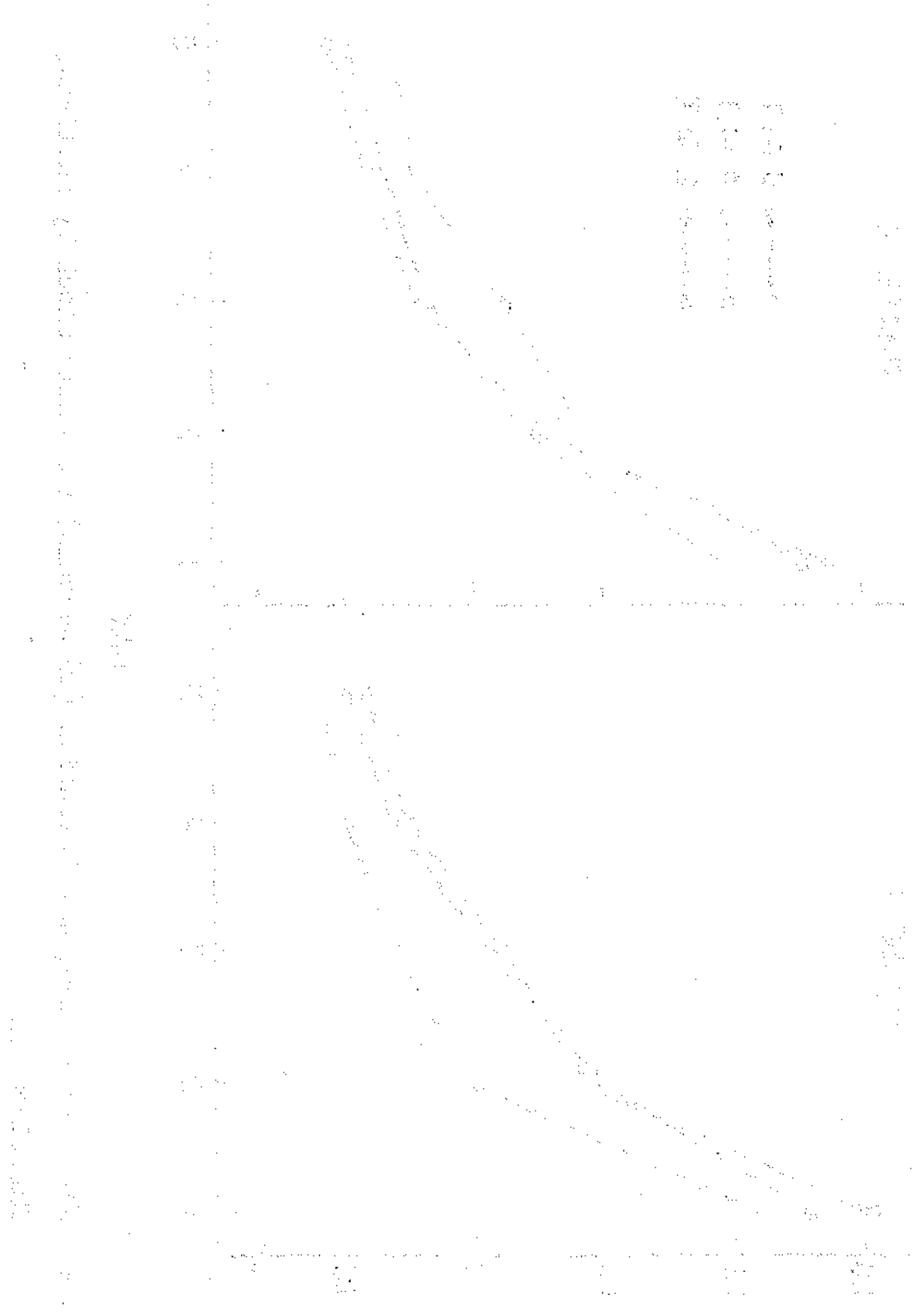


Figure 2. Mean correct response latencies (in seconds) per daily block by sublists, Experiment 1.



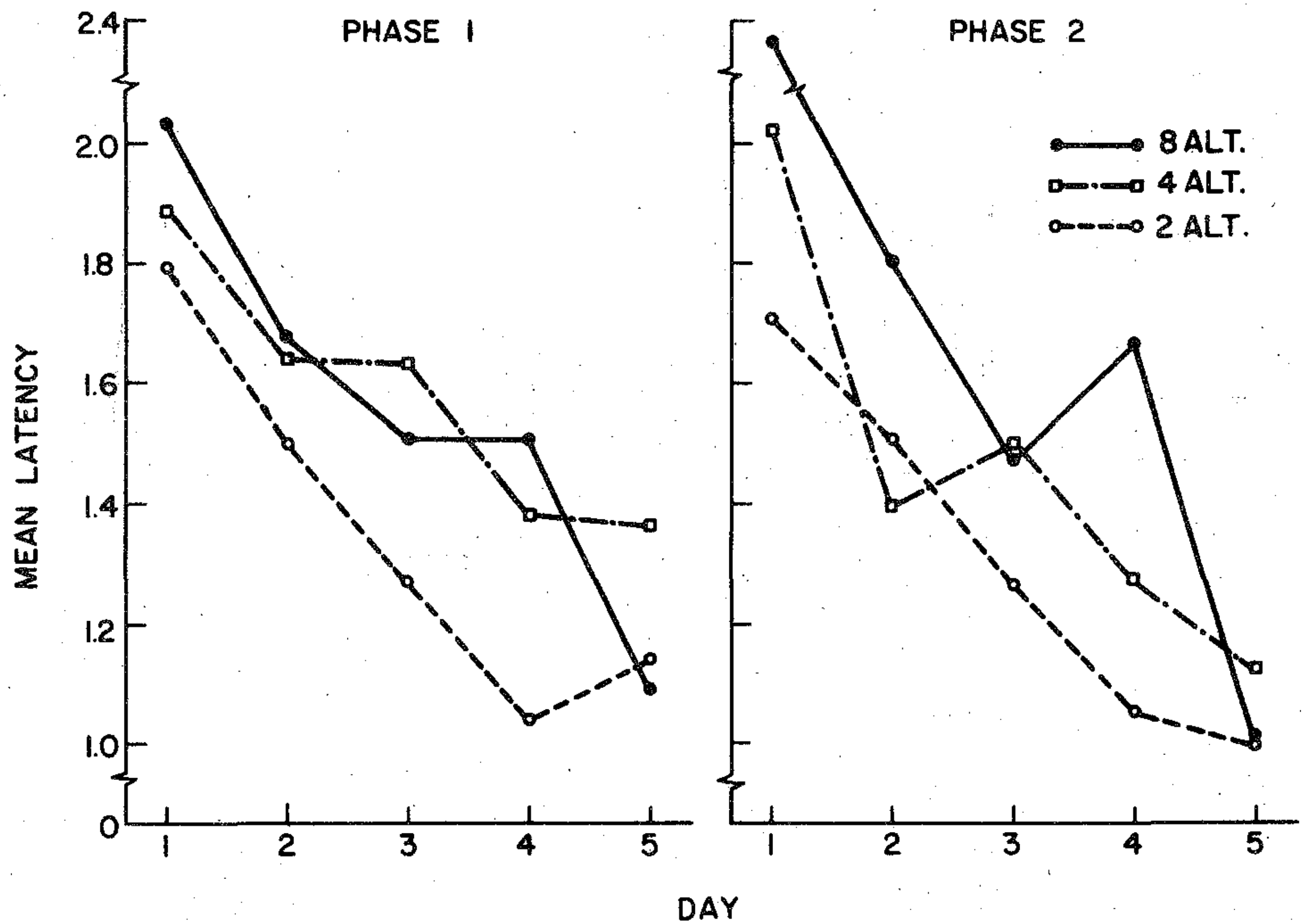


Figure 3. Mean error latencies (in seconds) per daily block by sublists, Experiment 1.

learning-to-learn effect. However, these curves cannot be taken as direct indicators of associative learning rates since learning of the response sets belonging to the three sublists during the first phase would be expected to modify the probabilities of correct guesses on unlearned items at the start of the second phase.

In order to draw conclusions regarding relative rates of learning, we can make use of the fact that paired-associate learning curves are well described by the function, common to both the linear and one-element models, among others (see e.g., Atkinson, Bower, and Crothers, 1965, pp 84-116)

$$P_n(C) = 1-[1-P_1(C)]\alpha^{n-1}, \qquad (1)$$

where $P_n(C)$ denotes the proportion of correct responses in the n^{th} block of trials and α is a constant with a value between 0 and 1. For the present situation, in terms of the one-element model, α represents the quantity $(1-c)^{12}$, where c is the probability that the item is learned on any one trial, and the exponent is 12 since there are 12 trials on each item per daily block. It can readily be shown that the quantity $\frac{5[1-P_1(C)]}{T_E}$, where $1-P_1(C)$ is the observed proportion of errors per S-item on Day 1 and T_E is the mean total number of errors per S-item over all blocks on a given list, provides an estimate of $\frac{1-\alpha}{1-\alpha^5}$, which can be taken as an index of learning rate that is independent of the initial probabilities of successes by guessing. This index can vary between 0 and 1, with larger values signifying higher relative rates of approach of the learning function to its asymptote. In the present data, the values of this index for the pooled sublists of Lists A and B, respectively, are .60 and .56. Thus there is evidently no appreciable change in learning rate from the first to the second phase when allowance is made for the change in initial values. When estimates of the index are computed for sublists with the same number of response alternatives, pooled over Lists A and B, nearly equal values of .56 and .54 are obtained for Sublists 2 and 4, with a substantially higher value, .65, for Sublist 8.

The learning curves for correct response latencies (Figure 2) exhibit a systematic change over trials similar to that of the frequency curves. If the mean latencies are converted to reciprocals, the response speed curves line up in the same order as the frequency curves initially, exhibit the same cross-over of the Sublist 2 and Sublist 4 curves, and differ from the frequency curves only in rising somewhat less steeply over the early trials.

Of even more immediate theoretical significance, perhaps, is the reproduction of the same pattern, in all essentials, during the learning of List B. This result is brought out still more clearly in Table 1, where mean correct and error latencies for the pooled sublists of each phase are compared. The facts that the List B latency curves start at the same level as those of List A and exhibit a similar course of decrement over trials seem difficult to reconcile with any interpretation except that the changes in latency reflect the course of associative learning.

Table 1

Mean Correct Response and Error Latencies per Daily Block,

Pooled over Sublists, Experiment 1

Day	Phase 1		Phase 2	
	Correct	Error	Correct	Error
1	1.73	1.92	1.68	2.09
2	1.32	1.62	1.32	1.54
3	1.20	1.49	1.19	1.42
4	1.06	1.33	1.08	1.30
5	1.00	1.23	1.02	1.06

Although the curves for error latencies are based on fewer observations than those for correct latencies, the set of curves presented in Figure 3 appears sufficiently orderly to leave little doubt about the overall pattern. Most importantly, it is clear that the changes in probability of correct responding during learning are not paralleled by a drawing apart of correct and error latencies, as might be anticipated on the basis of interpretations derived from Hull's theory (Hull, 1943; Spence, 1951) or from a combination of associative strength and signal detectability theory (e.g., that of Wickelgren and Norman, 1966) though it should be emphasized that these authors have not explicitly applied their models to recall latencies.

The sets of latency curves, for correct responses and errors, taken together, suggest that errors occurring early in learning arise from items still in an unlearned state, whereas errors occurring late in learning represent stimulus or response confusions among items which are in a learned state. The latency distribution associated with the learned state must have a lower mean than that of the unlearned state (in agreement with the assumption of Suppes, Groen, and Schlag-Rey, 1965), but within either state, correct response and error latencies come from a common distribution. With regard to the latter conclusion, the similarity of correct and error latencies in the terminal state of learning is apparent from a comparison of the corresponding points for Day 5 in Figures 2 and 3. To obtain evidence concerning the unlearned state, latencies have been analyzed separately for the precriterion portions of all protocols meeting a criterion of eight successive correct responses (286 out of 288 cases in List A and 280 out of 288 in List B). Comparison of the mean precriterion correct and error latencies for each list and sublist, presented in Table 2, supports the assumption of a common distribution in the case of Sublists 2 and 4, but reveals a consistent discrepancy for Sublist 8. Why the precriterion correct latencies should be somewhat lower than error latencies for Sublist 8 only is not readily explicable, but may be related to the fact that precriterion stationarity of error probability is generally realized closely only for two response alternatives, with departures increasing as the number of alternatives increases.

Before considering further the dependence of performance upon number of response alternatives, we should determine the extent to which the sublist response structure was learned during training on List A. The relevant data, in terms of initial proportions of correct responses on List B on each sublist, are presented in Table 3. The overall impression is that subjects had made substantial progress toward learning the response sets associated with each sublist, but that this learning was by no means complete. The marginal proportions of occurrence of the high, medium, and low frequency response alternatives (as defined above) are quite close to those expected if the subjects were restricting their guesses to the appropriate response subsets for items in each sublist. However, this correspondence may be somewhat coincidental. If the subjects had not actually learned the sublist structure, but had come to use the various digits, in guessing on unlearned items, with probabilities corresponding to the overall relative

Table 2

Mean Precriterion Latencies[a],

Experiment 1

Sublist Condition	Phase 1 Correct	Phase 1 Error	Phase 2 Correct	Phase 2 Error
8	1.68	1.95	1.72	2.23
4	1.80	1.81	1.83	1.75
2	1.79	1.69	1.47	1.53

[a]All means are based on at least 400 scores.

Table 3

Proportion of Correct Responses on Trial 1 of Phase 2 by Sublists in Relation to Overall Response Frequency, Experiment 1

Sublist Condition	Observed Response Frequency			Expected[a] Response Frequency		
	High	Medium	Low	High	Medium	Low
8	.427	.208	.365	.25	.25	.5
4	.563	.292	.146	.5	.5	0
2	.688	.146	.167	1	0	0
Mean	.559	.215	.226	.583	.250	.167

[a]Computed on the assumption that subjects had learned the assignments of response sets to sublists.

frequencies of occurrence of the various digits during training on List A, the expected marginal proportions for high, medium, and low frequency responses would be the same as those predicted on the assumption of sub-list learning and given on the right side of Table 3. The data for individual subjects exhibit wide differences. Four subjects seemed clearly to be guessing from the appropriate response set of each sublist; four exhibited relatively uniform distributions of trial 1 guesses over the full set of eight digits for all sublists; the remaining four appeared to use digits in rough proportions to overall frequency but without much evidence of distinguishing sublist assignments.

Regardless of the extent to which the subjects had mastered the sublist structures, in the sense of being able to transfer appropriate guessing strategies to new lists, it is clear that the course of learning was systematically and reproducibly influenced by the number of different response alternatives per sublist. The sublist learning functions, not only for correct response frequencies but for both correct response and error latencies, all exhibit the same trend, lining up in order of the number of response alternatives on the early days of each phase, but with the Sublist 8 functions gaining on the others. In most cases the Sublist 8 functions cross the others by the end of five days, so that terminal performance in each phase on each measure is characterized by a nonmonotone function of number of response alternatives. Mean values for trial of the last error were 14.12, 20.74, and 16.38 for Sublists 2, 4, and 8, respectively, in Phase 1, and 11.02, 12.32, and 10.95 in Phase 2. In order to exclude the possibility that these relationships result from the differential representation of high, medium, and low frequency responses in the three sublists, similar comparisons have been made using only data for the two responses common to all three sublists (e.g., digits 1 and 6 in the example given under Method). Although these functions are less stable, owing to smaller N's, the pattern of relationships among sublists is unchanged.

Experiment 2

In view of the results of Experiment 1, it becomes clear that questions concerning the relationship between response time and number of response alternatives will be more difficult to answer than one might have anticipated. Evidently differences in number of possible response alternatives per stimulus are inextricably associated with differences in rate of learning, and the pattern of findings suggests that the relatively small differences in mean response time for sublists having different numbers of response alternatives are a function of the differences in learning rate rather than the number of alternatives per se. Since maximum learning rates are evidently attained when the number of different responses in the list is the same as the number of stimuli, it seemed worthwhile to go on and determine whether there are appreciable asymptotic differences in response time as a function of number of items when number of stimuli and number of response alternatives vary concomitantly. This experiment was planned to assess response times for a relatively short and a relatively long list of items at the end of an extended period of training, with, in each case, a one-to-one assignment of responses to stimuli.

Design. Each subject learned two lists of paired-associate items, one list of 6 items and a second list of 24 items. Stimuli were four letter words with low emotional loading. All words started and ended with consonants. The same set of 30 stimulus words was used for all subjects. From this set of 30 words a unique random assignment to the 6-and 24-item lists was made for each subject.

Responses were two digit numbers. Each subject had 30 different responses chosen from a table of random numbers. A different set of responses was chosen for each subject. Numbers of the forms 0x, x0, xx, x(x+1) were not used.

Each subject was given 72 trials on each list over a period of nine consecutive weekdays distributed as follows:

Day 1	24 trials on the 6-item list
Day 2	12 trials on the 24-item list
Day 3	12 trials on the 24-item list
Days 4-9	8 trials per day on each list.

Each trial was a random permutation of the list. There was no indication to the subject when one trial ended and the next began.

On Day 10 subjects were given a choice reaction time test. Just as on Days 4 through 9, subjects received eight trials on each list. However, instead of a given stimulus word, the response number for that word was shown in the stimulus panel. The subject had only to press the corresponding buttons. He thus produced a series of responses completely analogous to those of the previous days but he was not required to remember any word-number associations.

Subjects. Six Stanford students were used as subjects. They were each paid $20 at the end of the experiment.

Apparatus. The response panel used in Experiment 1 was turned 90° so that the rows of buttons were now two columns of buttons numbered from 0 (nearest to subject) to 9. The numbers on the buttons were reoriented appropriately. The red button, now 2 in. to the right of the right-hand column, was used as a "hand rest" button. A stimulus word could not appear on the stimulus screen unless this button was held down by the subject. Once the stimulus word appeared on the screen the button could be released without affecting the rest of the cycle. A light inside the button remained on except while the button was held down.

A stimulus-response cycle proceeded as in Experiment 1 except that the subject, using the same finger with which he had been holding the hand rest button, produced a two digit response, first pressing one button in the left column, then one button in the right column. Buttons in the left column stopped the clock. Buttons in the right column triggered the display on the response screen and started the timers which controlled the rest of the cycle. Except for this use of two digit responses and the hand rest button the apparatus was the same as in Experiment 1.

Procedure. On the first day of the experiment the subject was seated at the response panel and read the instructions, essentials of which were as follows:

"You are to learn to associate two-digit numbers with one syllable words. A word will be shown on the upper screen. You are to make your best guess as to the number that goes with the word and then press the buttons that correspond to that number. The left column of buttons represents the left digit and the right column represents the right digit. Obviously you cannot know the correct answers at first.

"As soon as you press any buttons they will light up and the correct number will appear on the lower screen. The word will remain on the upper screen so that you can study the word and number together. After a short time both screens will go blank briefly before the next word is shown. Remember that the number on the lower screen is always the correct number no matter which buttons you push and that you must push both buttons before the correct number will appear.

"The procedure for responding is very important. Note the red button at the right of the response panel. This button must be held down at all times except when you are actually pressing the response buttons. Please use only the right hand and always use the same one finger for pressing the numbered buttons that you use to hold down the red button. When you press the response buttons be sure always to press a button in the left column before pressing a button in the right column. Always press one button in each column.

"We are interested in both the accuracy and the speed of your responses. Of course you will only be guessing at first but as you learn the correct numbers please respond as rapidly as you can without making mistakes. When you do not know the correct number please make the best guess you can."

Questions were answered by repeating or paraphrasing the appropriate sections of the instructions. The subjects were then given twenty-four trials on the 6-item list.

The subjects were given twelve trials on the 24-item list on Day 2, and twelve more trials on Day 3.

On Days 4 through 9 the subjects were given eight trials per day on each list. All trials for one list were completed, then there was a short break before the other list was presented. The order of presentation of the two lists was alternated from day to day with half of the subjects seeing the 6-item list first and half seeing the 24 item list first each day. The following instructions were read on Day 4:

"During the past three sessions you have seen two different lists. By now you should know them both well enough to respond without making mistakes. Now we are interested in finding out how rapidly you can respond, still without making errors.

"Today you will see both lists, each one several times. There will be a break between the two lists. Remember to use only one finger for pressing the buttons so that your response times will be comparable. If you should make an error just continue with the task."

On Days 5 through 9 the subjects were encouraged to respond as quickly and accurately as possible. Each session took about 35 minutes.

On Day 10 the subjects were read the following instructions:

"The procedure today will be a little different. Instead of words appearing on the upper screen you will see numbers on the upper screen. The numbers will be the same ones that you have been using as responses for the past nine days. When a number appears you are to press the response buttons that correspond to that number. Nothing will ever appear on the lower screen. We wish to see how fast you can respond when you don't have to remember the numbers.

"The procedure for responding should be just the same as you have used in the past. Use only one finger and hold the red button down except when responding. Just as in the past there will be a break between the numbers which represent the two lists."

The procedure was unaltered from the previous days except that the two digit response numbers replaced their corresponding words on the stimulus presentation screen.

Results

With regard to frequency of error scores, the learning curve for List 6 reached an asymptotic level of about 3 percent errors on Day 1 and that for List 4 on Day 2. Comparisons between the two are of no special interest since for all subjects List 6 was learned first. Except for the first few trials of the first few succeeding days, both functions remain at this common asymptotic level throughout the rest of the experiment.

With a criterion of eight successive correct responses, post-criterion latency curves for both correct and incorrect responses have been averaged by four trial blocks and these functions are presented for Lists 6 and 24 in the upper and lower panels of Figures 4 and 5, respectively. Both the correct and error latency curves decline substantially following the frequency criterion, with the error curves being higher on early days but approaching the same asymptote as the correct response curves, so that beyond Day 6 no appreciable differences remain.

The visual impression of near equality of asymptotic response times is documented in Table 4, in which mean correct latencies are presented for the two lists on the final day of paired-associate training, Day 9, and on the following day when both lists were given with the reaction time procedure. On both days differences between Lists 6 and 24 are statistically insignificant and small in magnitude relative to the means. Whether or not the null hypothesis is accepted, there would seem to be no positive support here for an assumption that recall responding involves a search process in which the set of available response alternatives must be scanned upon presentation of each stimulus.

Precriterion correct response and error latencies are presented by four trial blocks in Table 5 for the first four blocks, beyond which N's become too small for stability. Except perhaps for the drop from the first to the second block for the List 24 correct latencies, the values are generally relatively constant over blocks, with the correct latencies uniformly lower than the error latencies although, again, the difference is small in relation to the means. If the protocols are divided into subsets according to the trial of the last error, mean latencies for the subsets increase uniformly from trial 1 to the vicinity of the trial of the last error (exactly for TLE = 2, 3, 4

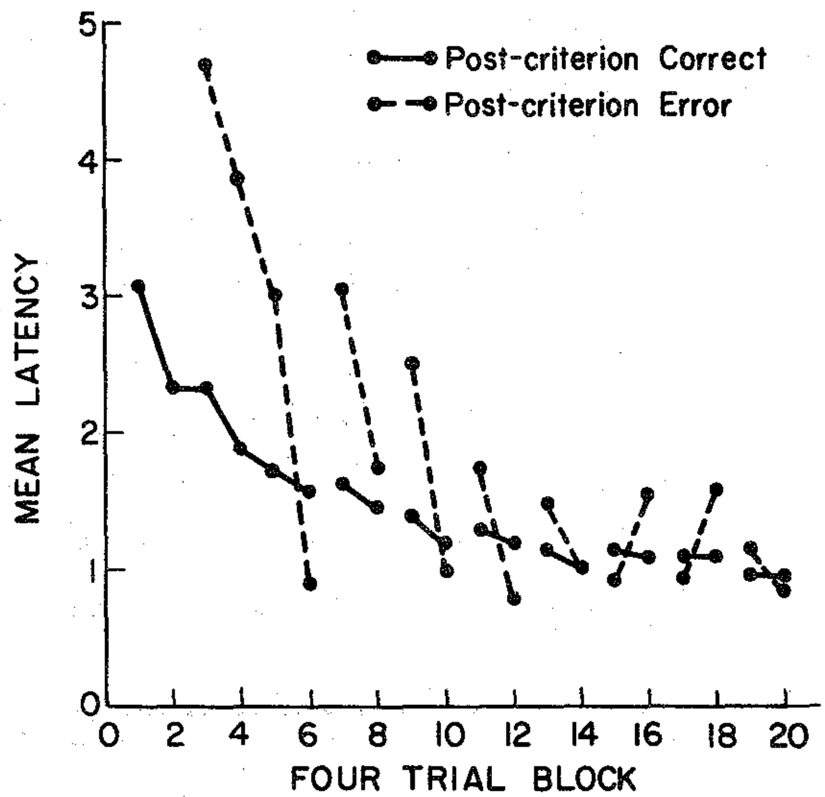


Figure 4. Mean postcriterion correct response and error latencies for the 6-item list, Experiment 2.

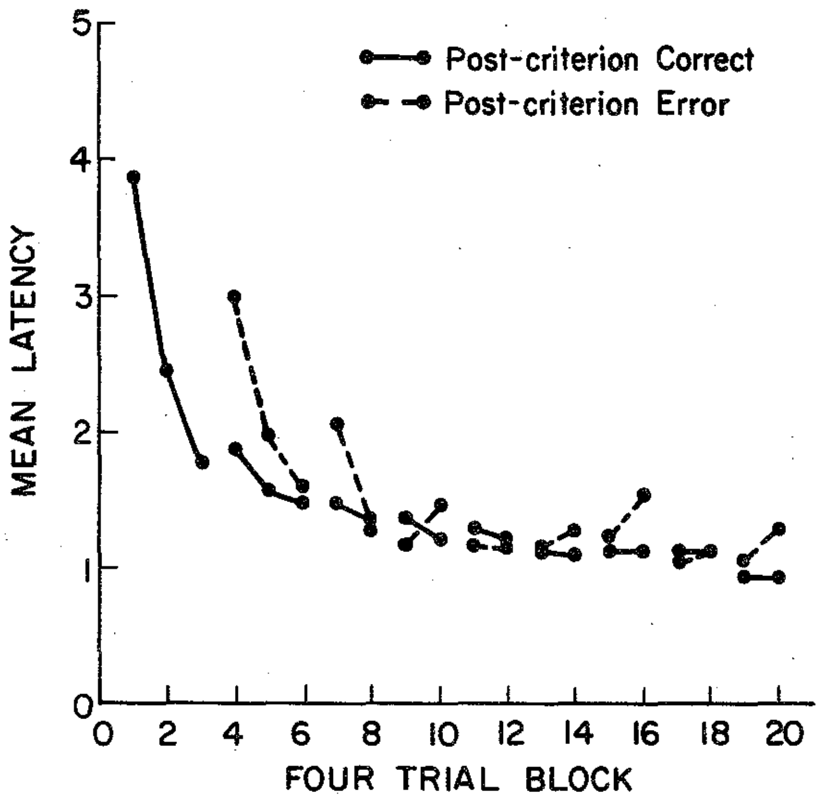


Figure 5. Mean postcriterion correct response and error latencies for the 24-item list, Experiment 2.

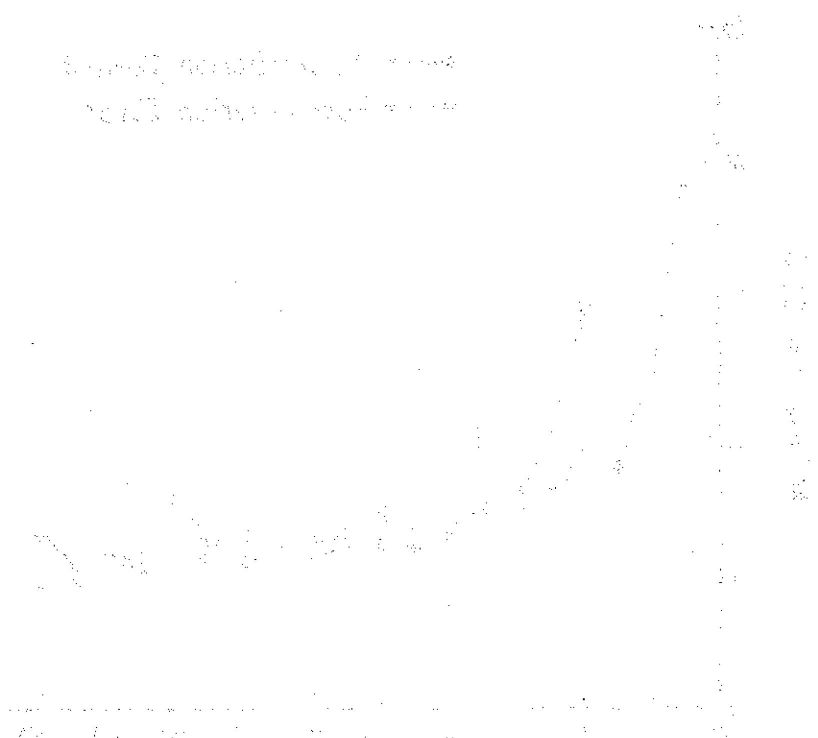


Table 4

Correct Response Latencies on Final Day of Paired Associate Training and on Reaction Time Test, Experiment 2

	Paired Associate		Reaction Time	
Subject	List 6	List 24	List 6	List 24
1	1.63	1.71	1.25	1.15
2	.97	1.01	.83	.96
3	1.34	1.47	1.04	1.05
4	.79	.86	.78	.85
5	.88	.87	.81	.83
6	1.00	.99	.86	.92
$\bar{X}$	1.11	1.15	.93	.96
σ	.38	.50	.26	.30

Table 5

Mean Precriterion Latencies by Four Trial Blocks, Experiment 2

	List 6				List 24			
	Correct		Error		Correct		Error	
Trials	Freq.	Lat.	Freq.	Lat.	Freq.	Lat.	Freq.	Lat.
1- 4	19	4.21	113	4.77	28	3.88	487	3.85
5- 8	29	4.03	81	5.12	57	2.88	219	3.37
9-12	19	3.42	30	4.72	78	2.62	96	3.75
13-16	11	4.70	5	5.18	43	2.82	54	3.26

and the preceding trial for TLE = 5, 6, beyond which N's become too small for analysis). The mean increase from trial 1 to the maximum is of the order of 2 sec. and the maximum value 4-5 sec. Beyond the trial of the last error, mean subset latencies in every case decline smoothly over trials to a common terminal level of about 1.1-1.2 sec., as seen in Figures 4 and 5.

Discussion

From the standpoint of our primary theoretical interest, the most important finding in the present study has to do with the similarity in the course of learning between Phase 1 and Phase 2 of Experiment 1 for both frequency and latency measures. The frequency curves (Figure 1) declined much more rapidly in Phase 2, as has frequently been observed when the same subjects learn successive lists, and which has been attributed to a "learning-to-learn" effect. However, our analyses indicated that the change from Phase 1 to Phase 2 could be accounted for in terms of increased probabilities of successes by guessing on unlearned items; with a suitable correction for this factor, the slopes of the learning curves did not differ appreciably between the two phases.

The similarity of learning curves for both correct and error latencies from Phase 1 to Phase 2 was even more striking, in each case the functions returning at the start of Phase 2 to the initial level of Phase 1 and then declining over a very similar course to essentially the same terminal levels as in Phase 1. Since the same response set was used in Phases 1 and 2, the reproducibility of learning functions from phase to phase must mean that the primary source of changes in latency in this situation was the learning of stimulus-response associations. Taken by themselves, the latency curves for Phase 1 might have been taken to reflect primarily adjustment to the experimental situation, changes in response availability or the like; however, one can scarcely imagine that any processes of that sort would be reset at the beginning of Phase 2 when the same subjects received continued training in the same experimental situation and with the same response set, but with new stimulus-response assignments.

Finer details of the pattern of frequency and latency curves support the conclusion of an intimate connection between latency changes and associative learning. On the early blocks of each phase, the error frequency curves tend to line up in the order of the number of response alternatives, with most errors being made on the 8 alternative sublists and fewest on the 2 alternative sublists; however over successive blocks in each phase the Sublist 8 curve gains at the expense of the others and reaches the lowest terminal value. Essentially the same patterns were observed for both correct and error latency curves. Initially the same ordering in relation to number of response alternatives is observed in all cases, except for one inversion in the case of correct latencies in Phase 1; and in all cases the Sublist 8 curve tends to gain at the expense of the others, yielding a terminal nonmonotone order with Sublist 4 curves ending highest and Sublist 8 curves, on the average, lowest.

A major question raised by our data is why the correct and error latency curves should change systematically in parallel fashion, with precriterion levels showing only a small advantage for correct over error latencies and with both functions declining over trial blocks to common asymptotes under all conditions in both experiments. This picture is scarcely what one would expect if latency were a measure of associative strength and if the effect of training trials were to increase associative strength for correct stimulus response connections and weaken it for incorrect ones. The interpretation we suggest may be given in two parts. Firstly, we see nothing to weigh substantially against the assumption common to most current mathematical models for paired-associate learning (see e.g., Atkinson, Bower, and Crothers 1967 ch. 2; Rumelhart, 1967) to the effect that, in the early stages of learning a list, both correct responses and errors come from an unlearned state, in which both types of responses occur as a result of guessing and have a common latency distribution. Secondly, in view of the similar functions for correct and error latencies, it appears that late in a series both types of responses must occur on items which have entered into a learned state. It does not seem likely that all late errors can arise from lapses of attention, since if that were the case there would be no reason for the error latency curves on late blocks to approach the same terminal levels as the correct latency curves. It seems likely, rather, that some at least of the errors occurring late in a series arise from items on which learning is incomplete in the following sense. Suppose that two items have some element of their stimulus members in common, but that each also has unique elements. If the associations between all of the elements of one stimulus and its correct response have been learned, the common element will cause confusion errors to occur on the other item whenever it appears until the unique elements of that item become associated with its correct response. Thus, all responses which arise as a result of learned associations are assumed to occur with low latencies, but they may be correct or incorrect depending on the experimenter's classification.

The tendency of all of the learning curves of Experiment 1 to line up in order of the number of response alternatives on early trials, and to do so more consistently and markedly in Phase 2 than in Phase 1, doubtless reflects the differential probabilities of successes by guessing on unlearned items, which would be expected to develop to the extent that the subjects learn the response sets belonging to the various sublists before they learn all of the correct stimulus-response associations. This factor seems quite sufficient to account for initial ordering of the frequency curves, but not for the latency curves. Particularly in the case of the error latencies, the relationship between latency and number of response alternatives on Day 1 of each phase is very marked, and it would be unaccountable if all early errors arose from a "guessing state" in which all responses had the same latencies. Surely this relationship must arise, directly or indirectly, from the acquisition of partial information, i.e., the association of response subsets with the common stimulus elements of the different sublists. But why should latency decrease when the subject restricts his guessing to a smaller subset of responses? One might suspect that the basis lies in the confounding of sublist size with overall response frequency: the smaller the sublist, the more frequently the response members occur, on the average; and latency of any response might decrease as a function of frequency of occurrence. However, a reanalysis of the data of Figures 2 and 3, using only the items of each sublist which have the same response members as Sublist 2, reveals no change in the ordering of the sublist curves. Thus we are led to suggest that the direct relationship between number of response alternatives per sublist and mean latency on the early days of each phase reflects the different times required by the subject to scan the appropriate response subset. Since the differences in mean latency among sublists tend to disappear asymptotically, it appears, further, that this scanning process occurs only when the subject is confronted with an unlearned item and searches his memory for a response which might be recognized as correct.

The consistent "crossing phenomenon," that is, the tendency for the learning functions for sublists with the larger numbers of response alternatives to overtake and cross those for smaller numbers of alternatives, would be expected if the occurrence of an error on a given item on one trial increased the likelihood of learning that item either on the same or on subsequent trials. Such a relationship has been embodied in Hintzman's (1968) computer simulation model for paired associate list learning, and it is compatible with the substantial results of Izawa (1966) on the effects of test trials upon later training trials. Also, this notion fits in with the plausible assumption that the occurrence of an error would tend to increase the likelihood that the subject would rehearse the given association, thus increasing the probability of correct responding on later trials.

With respect to more formal theoretical interpretations, an account of some of the trends in our data could be obtained from the model for paired-associate latency recently proposed by Rumelhart (1967). Rumelhart assumes that states of learning and transitions between states are essentially as assumed in the models proposed by Atkinson and Crothers (1964), Bjork (1966), and Greeno (1964). All of these models were developed to handle frequency data; all assume that in the course of learning an item begins in an unlearned state in which responses occur through guessing, then passes through at least one intermediate state, usually characterized as a short-term memory state, and finally is absorbed in a final learned state in which responses are always correct. Rumelhart took the additional step of associating latency distributions with the states. In unlearned states all responses are assumed to have latencies drawn from a common distribution with a relatively high mean. In the intermediate, short-term state, and initially in the final, long-term state, latencies are drawn from a distribution with a relatively low mean. Rumelhart further assumes that once an item is in the long-term state, it moves to a latency distribution with a higher mean if the item has not been presented for a relatively long interval, but then with further repetitions the latency declines toward the level characterizing learned items which have recently occurred.

This model would account for the similar course of learning in terms of correct response latencies and frequencies during both phases of Experiment 1 and for the generally similar pattern of results for frequency and correct response latency curves under the various conditions. It would not, as it stands, account for the curves for error latencies, and would have no means of handling the pattern of relationships regarding latencies as a function of number of response alternatives.

We are not prepared to give a full mathematical specification of a stimulus sampling model for our data, but we can outline the form such a model might take, drawing on the basic notions presented by Estes (1959 a,b) and LaBerge (1959). A set of stimulus elements is assumed to be associated with the stimulus member of each paired-associate item, and these are sampled randomly from trial to trial and become conditioned to correct responses in accordance with the standard assumptions. The response selection process would be much as assumed by LaBerge (1959). On a given trial, after being exposed to the stimulus and drawing (perceiving) a sample of elements, the subject would be assumed to scan the elements singly, skipping over elements that were unconditioned or were ambiguously conditioned (that is, associated with two or more responses) until he reached a uniquely conditioned element, whereupon he would make the corresponding response overtly. If no uniquely conditioned element were present, the subject would complete the scanning of the sample and then guess at random among the available responses. This machinery would account, without additional assumptions, for the close similarity of frequency and latency curves, for the similar forms of correct and error latency curves, for the observed relationships between Phase 1 and Phase 2 of Experiment 1. However, in order to handle the observed pattern of relationships between latency and number of response alternatives, together with the crossing of sublist learning curves over trial blocks, the assumptions of the model would have to be augmented in some such fashion as suggested in the preceding discussion.

References

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

Atkinson, R. C., and Crothers, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. Journal of Mathematical Psychology, 1964, 1, 285-315.

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

Crothers, E. J., Suppes, P., and Weir, R. H. Latency phenomena in the prolonged learning of visual representations of Russian sounds. International Review of Applied Linguistics, 1964, 2, 205-217.

Eimas, P. D., and Zeaman, D. Response speed changes in an Estes' paired-associate "miniature" experiment. Journal of Verbal Learning and Verbal Behavior, 1963, 1, 384-388.

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

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. Pp. 9-52. (b)

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

Hintzman, D. L. Explorations with a discrimination net model for paired-associate learning. Journal of Mathematical Psychology, 1968, 5, 123-162.

Horowitz, L. M., Norman, S. A., and Day, R. S. Availability and associative symmetry. Psychological Review, 1966, 73, 1-15.

Hull, C. L. Principles of behavior. New York: Appleton-Century-Crofts, 1943.

Izawa, C. Reinforcement-test sequences in paired-associate learning. Psychological Reports, 1966, 18, 879-919.

Kintsch, W. Habituation of the GSR component of the orienting reflex during paired-associate learning before and after learning has taken place. Journal of Mathematical Psychology, 1965, 2, 332-341.

LaBerge, D. L. A model with neutral elements. In R. R. Bush and W. K. Estes (Eds.), Studies in mathematical learning theory. Stanford, California: Stanford University Press, 1959.

Millward, R. Latency in a modified paired-associate learning experiment. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 309-316.

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

Spence, K. W. Behavior theory and conditioning. New Haven: Yale University Press, 1956.

Suppes, P., Groen, G., and Schlag-Rey, M. A model for latencies in paired-associate learning. Journal of Mathematical Psychology, 1966, 3, 99-128.

Wickelgren, W. A., and Norman, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, *3*, 316-347.

Williams, J. P. A test of the all-or-none hypothesis for verbal learning. *Journal of Experimental Psychology*, 1962, *64*, 158-165.