

Cue discrimination and forgetting

WILLARD N. RUNQUIST, LOUISE RENNEY, and KATHLEEN BIEDSDORFF
University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Subjects learned paired associate lists in which word triads were paired with nouns to a partial criterion; they were then given both an immediate and a 1-week delayed recognition test for discrimination among the triads and a cued recall test. Recall after 1 week was slightly poorer than that occurring immediately, but it was unaffected by various patterns of element identity among the triads. Discrimination improved somewhat over the 1-week interval. Further experiments showed that the immediate test was responsible for both reducing forgetting and improving discrimination, but both effects were independent of identity structure. Forgetting cannot be attributed to loss of differentiation of cues.

The primary purpose of this research was to examine the role of discrimination among cues in producing forgetting. The design necessary to accomplish this is straightforward. Subjects learn paired associate lists varying in similarity among cues and then return to the laboratory 1 week later to be retested. The testing procedure consists of a test of associative recall and an independent assessment of discrimination among the items composing the stimulus set.

Several experiments from the University of Alberta laboratory have shown a relationship between the number and position of identical elements in the cue terms of paired associates and the difficulty of learning the associates (Runquist, 1972). The most common interpretation of these effects has been in terms of learning discriminations among cues (Anderson & Bower, 1973; Gibson, 1940). It is not clear, however, whether element identity (similarity) affects forgetting. The bulk of the evidence has denied any effect (Underwood, 1961). However, Joinson and Runquist (1968) found that similarity retarded forgetting, and Battig (1966) has persistently maintained that interference during acquisition should facilitate long-term retention.

Irrespective of the effect of similarity on forgetting, it is important to determine whether discrimination among cues, once established, is retained over long time intervals or whether it shows the forgetting characteristic of associative information. We might add that forgetting is not a foregone conclusion. Ellis and Daniel (1971) have shown that the ability to distinguish new nonsense shapes from well-studied prototypes does not deteriorate over a 28-day interval. It is possible that intralist discriminations show the same stability.

EXPERIMENT 1

Method

The experiment consisted of five parts and was identical for all subjects. There were four different conditions, but these were

This work was supported by Grant A088 from the National Research Council of Canada. James Tousignant, Caroline Renney, and Carrie MacWilliams aided in the collection of the data.

defined only by the structure of the list learned. The first part of the experiment was a practice task, the purpose of which was to familiarize the subject with the testing procedure. Subjects were given study-test sequences on six pairs composed of a geometrical form and a two-digit number. Pairs were presented singly at a 2-sec rate on study sequences, and the geometrical form was presented alone at the same rate on test sequences. Three study-test sequences were given, with the subject attempting recall of the numbers on the test sequence. Immediately after this learning phase, a testing phase was begun. In the test cycle, five of the six forms were presented at a 2-sec rate and the subject attempted to recall the numbers. Then a set of question marks appeared for 2 sec, followed by either the sixth form or one of the five already presented on that cycle. The subject first rated the form from 1 to 6, with 1 indicating certainty that it had appeared on that test cycle and 6 indicating certainty that it had not. After the rating had been given, the subject attempted to recall the correct numbers. This "probe" remained visible for 10 sec, after which a set of asterisks appeared, indicating the start of another test cycle. Four test cycles were given to all subjects, with two probes being members of the test cycle and two being the missing member. If the subject still did not appear to understand the procedure, four additional cycles were given, along with further information.

The second part of the experiment was identical to the learning phase of the practice list, except that study-test sequences were continued until the subject correctly recalled at least four items on a given test sequence. The pairs were composed of three consonant-vowel-consonant words as cue items and two-syllable nouns as target items.

The third part of the experiment consisted of 12 test cycles. In each cycle, five cue items were presented for recall, followed by question marks and then either a repeat or a missing probe. Each cue item was omitted twice during the 12 cycles, and each stimulus was probed once as a repeated item and once as a missing item. The testing sequence was the same for all subjects.

The fourth part of the experiment began immediately after the 12 test cycles. The subject was given a sheet of paper containing 15 numerical series problems and was allowed to work on these problems for 2 min. The purpose of this task was to minimize immediate rehearsal of the material.

The fifth part of the experiment occurred 1 week later. The subject was telephoned the night before and reminded of the appointment. The return test consisted of an exact repetition of the third part of the experiment, the 12-cycle combined recognition-recall test.

Four different lists defined the conditions of the experiment. The same six words were used as target items in all lists. The cue triads for the low-similarity (LS) lists were constructed from 18 different words. For each of the other lists, only eight different words were used. In one list, the position rule (P) list, the second word in each triad was unique to that triad and

the remaining two words were identical in all triads. The position of these words was reversed for three triads. In another list, there was also a unique word in each triad, but its position varied from triad to triad, so that it was in the first position in two triads, the second position in two triads, and the third position in two triads. This list defined the exclusion rule (E) condition, since the discriminative element could be defined by excluding the two redundant elements irrespective of position. The final list may be called the multiple redundancy (M) list, as there was no single word that was not duplicated in some other triad. If the letters A-H stand for the eight elements, the structure of the six triads may be described as ABC, DAC, CBE, FEG, FAH, and DHG. Since the E and M lists were more difficult than the LS and P lists, one extra study-test sequence was given on the former lists after the subjects had attained criterion in an attempt to equate recall performance on the immediate test sequence.

The subjects consisted of 80 introductory psychology students who participated as part of a course laboratory requirement. All subjects made appointments for two sessions 1 week apart, and then they were assigned to conditions as they appeared at the laboratory according to a prearranged scheme that randomized the order of conditions within blocks containing each condition once. Nine additional subjects were tested or partially tested, but their data were discarded because of experimenter error or equipment failure. Subjects were told nothing about the second session.

Results and Discussion

All statistical comparisons use the 5% rejection region.

Acquisition. The mean numbers of study-test sequences to reach criterion for the LS, P, E, and M conditions were 5.05, 5.40, 8.20, and 14.60, respectively. Analysis of variance produced a significant F value [$F(3,76) = 16.56$, $MSe = 23.62$]. The corresponding mean total errors to criterion were 19.35, 22.25, 34.40, and 65.00 [$F(3,76) = 16.03$, $MSe = 543.57$]. These differences are comparable to those obtained in other experiments using the same structures (Runquist & Sekulich, 1979).

Recall. Table 1 presents the recall data for both immediate test cycles and those given 1 week later. The mean total recall performance represents the combined results of the 12 paced recall tests and is based on 60 opportunities/subject. The recall to the probe stimulus was essentially unpaced and was based on 12 opportunities/subject. The values in the table are the means of percentages obtained from each subject.

The criterion procedure was designed to produce equal performance for the four groups on immediate recall whether paced or not; that is, the extra trial

Table 2
Mean Correct on Delayed Recall for Items Varying in Number Correct on Immediate Recall: Experiment 1

	List Structure							
	LS		P		E		M	
	Mean	N	Mean	N	Mean	N	Mean	N
0-3	.7	17	1.2	18	.5	17	1.3	19
4-7	3.0	11	4.4	9	4.7	11	4.3	16
8-10	7.6	19	7.1	19	8.4	19	7.6	15

Note—N = number of subjects contributing items to that category.

given the E and M groups should have compensated for the greater postcriterial drop in these conditions. Clearly, however, the procedure had limited success. While performance was equivalent in the LS, P, and E conditions, recall in the M group remained at a lower level [when the M condition was included in the analysis, $F(3,76) = 6.47$ for paced recall and 4.73 for probe recall; when it was not, $F < 1.00$ for both measures]. Direct comparisons are hence not meaningful when the M condition is included, and we shall deal with this group separately in most analyses described below.

An overall analysis of the two sets of test cycles (excluding M) showed that significant forgetting did take place over the 1-week retention interval [$F(1,56) = 44.81$ for paced recall and 7.45 for probe recall]. The critical interaction between similarity and retention interval, which could indicate differential forgetting for the three lists, produced $F < 1.00$ in both cases, despite the fact that the E group appeared to show less forgetting.

In order to compare forgetting in the M group with forgetting in the others, individual items for which performance was identical on the immediate test were tabulated for the various groups. On the 12 trials of the immediate paced test, each item was tested 10 times. Items can thus be classified by the number of times they were correctly recalled on the immediate test, and their fate on the delayed test can be recorded. Although the M condition, because of poorer immediate performance, would have a different distribution of items than would the other groups, the various conditions may be compared for each class of items.

In actually performing this analysis, three categories were used: correct 0-3 times, correct 4-7 times, and correct 8-10 times. For each subject, the mean number of items correctly recalled on the delayed test was determined for each category. These results are shown in Table 2. The sample size varies because some subjects did not contribute items in some categories. Generally, however, there were no systematic differences among the groups. While the shifting samples precluded an overall statistical test, none of the differences between groups was significant; the largest t was 1.31 between LS and E for the strongest items. More to the point, however, is the fact that recall for the M group was quite in line with that in other conditions. We thus

Table 1
Mean Percent Recall: Experiment 1

	List Structure				s_m
	LS	P	E	M	
Mean Paced Recall					
Immediate	68	65	68	44	4.0
1 Week	53	50	59	36	5.0
Mean Probed Recall					
Immediate	62	63	53	37	5.6
1 Week	51	51	52	31	5.8

have no evidence that structural similarity among cues affected forgetting rate.

Recognition (discrimination). The subject's ability to determine whether the probe stimulus was or was not included in the test cycle was used as an index for cue discriminability. There are two ways that the test may be conceptualized, each of which leads to a somewhat different approach to the data.

The most obvious analysis is to view the test as a yes-no recognition test. The paced test cycle provides the subject with the set of "old" items. The probe may then be either old (repeat) or new (missing). By making usual assumptions of signal detection theory (Banks, 1970), it is possible to use that model for analyzing the recognition data. Although the appropriate measure of criterion-free discrimination is d' , we followed the practice used by Underwood (1974) and based our analysis on hits plus correct rejections (total correct classifications of repeat and missing probes), which correlates almost perfectly with d' .

The logic behind the alternative procedure is more tortuous (Runquist, 1974). We assume that on the paced test cycle, the subject codes each cue and uses it to guide retrieval. When the probe is presented, the subject also encodes that stimulus. If it is the missing item, a recognition error will result when the probe code is indistinguishable from that of one of the test cycle items. Thus, errors on missing items indicate the subjects' inability to discriminate among various cues. Performance on repeat cues is not so neatly interpreted. A subject may confuse the repeated probe with another item that was tested and still make the correct recognition decision. According to this rather simplistic model, then, performance on missing probes (false alarms) is the appropriate measure of discrimination among cues. We will examine the data from both perspectives.

We make no claim that recognition performance "measures" discrimination in any absolute sense, but only that certain differences in recognition performance reflect comparable differences in discrimination. Clearly, the judgment of a particular probe involves other factors (e.g., short-term memory for a recent item occurrence). The point is, however, that anything the subject learns about the cue items during associative learning may also be reflected in the recognition scores, and changes in the state of this knowledge could occur over time. It is these effects we label "discrimination effects" and are concerned with at this point.

Although subjects made confidence ratings about the cue items, we will not report these data. Few subjects used ratings other than the extremes, and by treating ratings of 1, 2, and 3 as "old" and 4, 5, and 6 as "new," the percentage of correct decisions captures any trends in the data revealed by rating measures. These percentages are shown in Table 3. The bottom two rows of the table provide the percentages of total correct decisions on both repeat and missing probes. Not only is there no evidence of forgetting, but also three of the four lists

Table 3
Mean Percent Correct Recognition Decisions: Experiment 1

		List Structure				
		LS	P	E	M	s_m
Repeat	Immediate	81	85	67	68	3.7
	Delay	81	83	80	74	3.5
Missing	Immediate	74	58	69	60	6.1
	Delay	80	58	83	68	6.2
Total	Immediate	78	72	68	64	4.1
	Delay	81	71	82	71	4.2

show improved performance on the delayed test. The main effect of retention interval was significant [$F(1,76) = 8.21$], as was the interaction of similarity structure with retention interval [$F(3,76) = 2.65$]. Overall, the improvement in discrimination performance was from 70% to 76%. Only the P structure failed to produce an increase, and the improvement was largest with the E list.

The main effect of similarity structure approached but did not attain the required level of significance [$F(3,76) = 2.43$, $p = .07$; with the M group excluded, however, $F = 1.53$, $p = .23$]. Differences among the LS, P, and E groups would be of considerable importance because recall differences among these groups did not occur. Thus, it would appear that when recall is equated, discrimination does not significantly differ.

Nevertheless, there was the indication of a difference among the similarity structures. Within the signal detection model, differences in the subjects' response criterion were reflected in an interaction between treatments and hits vs. correct rejections (repeat vs. missing items). Table 3 does indeed show such an interaction, and the analysis of variance verified its significance [$F(3,76) = 6.43$]. The interaction appeared to be due primarily to the adoption of a more lenient criterion in the P condition. While we have no reasonable psychological explanation for this result, it has been found in other experiments (Runquist, 1975b).

If only performance on missing items is used to index discrimination among cues, the conclusions are much the same. When all four groups were considered, similarity structure was marginally significant [$F(3,76) = 2.67$], even with the M group excluded [$F(2,57) = 3.33$]. This effect was largely due to fewer correct choices in the P condition. The main effect of retention interval was significant with all groups included [$F(1,76) = 5.32$], but not with M excluded [$F(1,57) = 3.45$, $p = .07$]. The interaction was not significant in either case [$F(3,76) = 1.22$, $F(2,57) = 1.31$].

While these analyses contain some ambiguities with respect to the effect of similarity structure, it is quite clear that discrimination performance, however measured, does not decrease over 1 week.

Discrimination and recall. Several analyses were conducted to determine whether discrimination performance was related to recall performance. The results

of these analyses were generally negative. Some of the more salient ones will be described.

Two aspects of the data presented so far bear upon this problem. On the immediate test, both paced and probe recall in the E and P conditions were as good as recall in the LS condition, yet discrimination was poorer for E on repeat probes and for P on missing probes. However one chooses to measure discrimination, equal recallability does not insure equal discriminability.

With respect to retention interval, recall decays over time, but discrimination does not. This result, which complements that obtained by Ellis and Daniel (1971), does not necessarily mean that discrimination plays no role in recall, but it does indicate clearly that forgetting does not result from a loss of differentiation among cues (Gibson, 1940).

It has become more or less common practice to examine the relationship of recall to other processes by conditionalizing recall on successful performance on the presumably related task. While the procedure is fraught with problems, particularly regression effects (Runquist, 1973), such results are sometimes illuminating. In the present case, we counted those probes that produced a correct recognition decision and then we determined the percentage of those that were also effective cues for recall. The values are presented in Table 4. A comparison of the percentages with those in Table 1 reveals that the conditional probabilities deviated little from the overall percentages, despite the fact that regression effects should have resulted in correctly discriminated stimuli being more effective recall cues than nondiscriminated stimuli.

Finally, we examined the product-moment correlation coefficients between discrimination and recall. They provide the only evidence of any relationship between the two processes. The correlations between total correct discriminations and total paced recalls were .34, .19, .65, and .68 for the LS, P, E, and M conditions, respectively. Only the latter two are significant, but it is in these two conditions that discrimination should be more important. We will not present other correlations, because they show essentially the same pattern but appear to be less reliable, since the individual scores were based on fewer tests. It must be remembered that there is a certain amount of ambiguity in the interpretation of these correlations. Empirically, the correlation states that subjects who do well on recall

also perform well on discrimination. It does not necessarily imply a causal relation between the two processes. The commonality could result from the fact that the two tasks require the same knowledge about cue differences, or from sharing of some common learning ability or motivational level. In the absence of any further positive evidence, we are inclined to treat the correlational data conservatively, albeit as positive evidence.

Correlations were also computed between various measures of discrimination and forgetting. None of these correlations approached significance.

EXPERIMENTS 2 AND 3

The purpose of these two experiments was to examine more closely two of the results obtained in Experiment 1. The first result was the increase in discrimination accuracy during the 1-week retention interval. The second was the lack of relationship between similarity structure and forgetting. Also of some interest was the correlation between discrimination and recall. In these experiments, we focused on the E condition. This choice was motivated by the fact that (1) the improvement in discrimination was largest in this condition, (2) the correlation between recall and discrimination was sizable in this group, (3) our procedure equated immediate recall for this group and the LS control, and (4) there was no evidence of criterion effects on the discrimination test.

Essentially, the experiments were designed to determine the extent to which the results of Experiment 1 depended on giving the subjects both immediate and delayed testing.

Method

With two exceptions, the procedure was identical to that in Experiment 1. The most important exception was the placing of the 12 paced recall-probe test cycles. In Experiment 2, only the E list was used and there were three different conditions. In one condition, Group E-D, the test sequence was given only once, 1 week after acquisition. A second group had both immediate and delayed tests, exactly as in Experiment 1, with the second test following the first by 1 week (Group E-ID). The third group received three test sequences: two immediately following acquisition and one 1 week later (Group E-2I).

In Experiment 3, only the LS list was used. There were two groups, one duplicating the procedure in Experiment 1 (both immediate and 1 week tests Group LS-ID) and one being tested 1 week after acquisition only (Group LS-D).

The second difference in procedure was the elimination of the practice task on nonsense shapes and numbers. While the complexity of the discrimination test sequence caused us some anxiety in this respect, it did not seem appropriate to have subjects practice on a task that in some cases was not going to be given until 1 week later.

The subjects in Experiment 2 were introductory psychology students selected and assigned to conditions as in Experiment 1. Experiment 3 was performed after Experiment 2 had been completed. As a result, about half of the subjects for Experiment 3 were paid volunteer summer session students and half were introductory psychology students obtained as above. There were 100 subjects altogether, 20 per group, plus 19 subjects whose data were either incomplete or discarded because of various failures.

Table 4
Recall Conditional on Discrimination: Experiment 1

	List Structure				s_m
	LS	P	E	M	
	Repeat Probes				
Immediate	65	62	59	38	5.5
Delayed	58	54	57	32	5.7
	Missing Probes				
Immediate	63	61	61	36	5.7
Delayed	51	50	53	30	5.8

Results

The mean number of trials to criterion for Groups E-D, E-ID, and E-2I were 10.0, 11.3, and 10.2, respectively. Corresponding mean total errors during acquisition were 42.9, 49.1, and 42.7. Differences among groups were not significant ($F < 1.00$). For the two LS groups, the mean trials to criterion were 6.0 and 5.4 for LS-D and LS-ID, respectively, and the corresponding mean error totals were 26.65 and 21.20. Neither difference was significant ($F < 1.00$). It may be noted that all groups learned somewhat more slowly than the subjects in Experiment 1.

The recall data from the test cycles are presented in Table 5. Although the table presents the results of both paced and probe recall, we shall consider only the former in detail. All of the effects present in paced recall were also present on recall to the probe, but the former data involved more observations and hence yielded effects that tend to be somewhat more reliable.

The most salient feature of these results is the drastic effect of the immediate test on forgetting. Delayed recall produced losses near 50% without the immediate test, but the test completely abolished forgetting. Comparison of the immediate and delayed recall for LS-ID and E-ID groups separately produced $F < 1.00$ in both cases. On the other hand, analysis involving the immediate test of the ID groups with the 1-week test of the D groups produced a significant main effect of retention interval [$F(1,76) = 61.78$]. Similarity structure was not significant [$F(1,76) = 1.51$], and the interaction indicative of differential forgetting also fell short of the critical region [$F(1,76) = 2.87, p = .094$].

Unfortunately, the question of differential forgetting must remain unsettled. Without the practice task, our procedure for equating the E and LS groups on immediate recall was not completely successful, as performance on the E list was lower than that on the LS list [$F(1,38) = 4.71$]. This difference could produce a bias favoring retention in the latter group, and while there are a number of ways of handling the problem (Underwood, 1964), none of them provides an unambiguous resolution with the present data. The most cogent analysis used the acquisition phase data to estimate immediate recall for each subject in the 1-week conditions, thus

enabling the computation of both absolute and percent loss for each subject (Richardson & Underwood, 1957). Using this procedure, the predicted immediate recall scores for the 12 test sequences were 40.2 for the LS group and 32.9 for the E group. The mean number of items forgotten was 22.5 for LS and 14.5 for E [$F(1,38) = 3.76, p = .06$]. For mean percent loss, the respective values were 56% and 44% [$F(1,38) = 1.04, p = .31$]. Thus, while we have more or less consistently found less forgetting when similarity is high, the effect is not reliable enough to overcome the high variability normally obtained with loss scores.

Multiple testing effects can be most clearly seen in comparisons among the 1-week tests for all groups [for the two LS groups, $F(1,38) = 28.00$, and for the E groups $F(2,57) = 15.71$]. Group E-2I also provides an interesting comparison in this respect. Recall on a second test sequence given immediately was significantly better than that on a second test sequence that was delayed by 1 week [$F(1,38) = 6.30$]. It is impossible to determine whether this result means that there were losses in associative information in Group E-I that were countered by improvement due to retesting or whether test-retest effects were simply more potent when practice was massed. These effects are particularly striking in light of the fact that each test consisted of 12 trials including 10 tests on each cue. One might expect testing effects would be at asymptote after 12 trials, but apparently they were not.

The results on the discrimination test are presented in Table 6. Regardless of whether total correct recognition decisions or only correct identification of missing items is taken as the measure of discrimination, the results are the same. A two-way comparison of the immediate tests for the LS-ID and E-ID groups with the 1-week test for the LS-D and E-D groups showed that retention interval was not significant [$F < 1.00$ for total correct, and $F(1,76) = 1.54$ for missing items]. The interaction of retention interval with similarity structure was not significant ($F < 1.00$ for both measures). The E groups performed more poorly than the LS groups on both total correct [$F(1,76) = 6.75$] and missing items [$F(1,76) = 5.36$].

This outcome, in conjunction with the results obtained on the 1-week test for LS-ID and E-ID, makes it quite clear that the improvement over time obtained in discrimination performance in Experiment 1 was, in fact, a test-retest effect. The improvement in total correct was significant for the E-ID condition [$F(1,57) = 9.79$] and for the LS-ID condition [$F(1,57) = 4.83$]. For missing-item recognition, the E-ID group again showed improved performance [$F(1,18) = 8.82$], and the LS-ID group fell just short of producing significant improvement [$F(1,18) = 3.65, p = .07$].

The final comparison is an assessment of the effect of delaying the second test. The results of the 1-week test for the E-ID group were compared with the second test for the E-2I group. Neither total correct nor missing-item recognition showed a significant effect.

Table 5
Mean Percent Items Recalled: Experiments 2 and 3

Test	List Structure					s_m
	LS-D	LS-ID	E-D	E-ID	E-2I	
Paced Recall						
Immediate 1		71		58	67	4.2
Immediate 2					76	4.5
1 Week	28	69	30	56	72	6.0
Probed Recall						
Immediate 1		70		51	59	5.5
Immediate 2					72	4.9
1 Week	23	68	23	58	70	6.0

Table 6
Mean Percent Correct on Discrimination Test:
Experiments 2 and 3

	List Structure					<i>s_m</i>
	LS-D	LS-ID	E-D	E-ID	E-2I	
			Repeat Tests			
Immediate 1		82		78	83	4.3
Immediate 2					88	2.7
1 Week	87	88	74	82	80	3.7
			Missing Tests			
Immediate 1		77		65	66	5.5
Immediate 2					82	4.5
1 Week	71	87	63	82	93	3.8
			Total			
Immediate 1		79		72	75	4.7
Immediate 2					85	3.5
1 Week	79	87	69	82	87	3.8

We have one other incidental finding to report. If the signal detection model is considered to be an appropriate analytic procedure for this paradigm, there is some evidence for the existence of criterion differences, although it is somewhat ambiguous. Interactions of repeat and missing-item recognition with similarity [$F(1,114) = 3.53, p = .06$] and immediate and 1-week test [$F(1,114) = 2.28, p = .13$], apparent in the data, fell short of significance, however.

With respect to the relation between recall and discrimination performance, the results in Experiments 2 and 3 are more convincingly negative. These results are of several sorts: (1) Recall deteriorated considerably over 1 week, but discrimination showed no change. (2) The recall of correctly discriminated items closely approximated that of overall recall. The measure did not eliminate or even reduce differences in recall. (3) Product-moment correlation coefficients based on combined immediate test for Groups E-ID and E-2I were .20 for total discrimination performance vs. paced recall and .22 for missing-item recognition vs. paced recall. These correlations are considerably lower than those obtained in Experiment 1, and they fell short of significance ($r = .26$), despite the fact that they were based on larger samples.

DISCUSSION

There are several salient conclusions to be drawn from the results of these experiments. In most cases, the data are unambiguous. The conclusions are as follows. (1) Discriminability among cues did not deteriorate over a 1-week time interval. (2) Discriminability was not related to either immediate recall or to forgetting rate. (3) Both recall and discrimination were facilitated by previous tests. The effect was time dependent for recall, but time independent for discrimination. That is, a delayed recall retest was influenced more by a previous test than was an immediate recall retest, but delay did not affect the amount of facilitation on the

discrimination test. Testing effects were independent of similarity among cues. (4) There is some evidence that similarity among cues retards forgetting, but the difference is eliminated if an immediate test is given. This conclusion is statistically ambiguous.

Of the first two conclusions, little need be said. Other studies (Nelson, Brooks, & Wheeler, 1975; Runquist, 1975a, 1978; Runquist & Sekulich, 1979) have suggested that cue discrimination is independent of recall. The present research verifies this proposition, in that recall deteriorates over time but discrimination does not. At the theoretical level, the results indicate that forgetting is not produced by loss of cue differentiation.

It would be tempting to argue that discriminative coding of this class of cues, once established, is resistant to forgetting, but we wish to remain properly conservative on this claim. This conclusion rests upon the assumption that performance on the immediate recognition test improves during associative learning. Since we do not have a measure of discrimination prior to learning, we cannot provide evidence that this is the case.

The issue is further clouded by the fact that studies designed to show learning effects in this and similar cue discrimination tests are ambiguous, with some showing substantial improvement with study (Runquist, 1975b, 1978) and some not (Nelson et al., 1975). Positive effects have generally been small. On the other hand, there is ample evidence that variations in similarity structure produce consistent differences in discrimination performance. In the present case, some of these differences were not obtained after learning to criterion. Thus, our argument that discrimination was established is indirect, but nevertheless cogent. Irrespective of this issue, the conclusion regarding the independence of the two measures and the processes they presumably represent holds.

The other results are related to testing effects. Of course, improvement of recall with successive tests is not a unique finding, although similar effects on discrimination tests have not been previously demonstrated. The fact that the test effects in recall seem to interact with retention interval does raise some interesting questions regarding the theoretical interpretation of forgetting phenomena. The failure of the discrimination task to show a similar dependence again argues for the independence of the knowledge necessary to perform each task.

The general notion that the subject establishes general or specific strategies for performing whatever retrieval operations are required is not uncommon in memory theory. It has been variously conceptualized and studied as learning to recall (Greenberg & Underwood, 1950), warm-up (Irion, 1949), the availability-accessibility distinction (Tulving & Pearlstone, 1966), retrieval strategies (Gotz & Jacoby, 1974), and retrieval plans (Greeno, James, & Da Polito, 1971). While there has been little discussion of forgetting as the loss of these skills, the point has occasionally been made that

performance on delayed tests may suffer if the appropriate operations are not applied after a retention interval (Irion, 1948)

Additional test trials immediately after acquisition may eliminate or delay forgetting by providing sufficient retrieval practice to forestall the loss of the appropriate skills. Allen, Mahler, and Estes (1969) not only demonstrate the result but also offer essentially the same interpretation. The general failure of most variables to affect forgetting rate (Postman, 1971) may be due to the fact that the loss of retrieval operations is independent of most variables that have been investigated, simply because these variables have profound effects on acquisition.

Just what these retrieval operations may be and why they are lost remains unspecified, and we are not prepared to speculate further at this time. Nevertheless, the fact that the effects of some variables depend upon prior testing (Tulving & Watkins, 1974) and that the effects of other variables disappear with repeated testing (Runquist & Snyder, 1969) underscores the possible role of these processes in determining forgetting.

Testing effects have not always been interpreted in terms of retrieval operations, and some investigators prefer to consider the phenomena in terms of additional study of to-be-recalled information (Thompson, Wenger, & Bartling, 1978). While we do not wish to deny the possibility of such a straightforward interpretation, we do not think that it can account for all of the test effects observed. In particular, we agree with Martin, Fleming, and Nally (1978) that "pure" study effects are more likely to exist in free recall procedures.

REFERENCES

- ALLEN, G. A., MAHLER, W. A., & ESTES, W. K. Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 463-470.
- ANDERSON, J. R., & BOWER, G. H. *Human associative memory*. New York: Winston-Wiley, 1973.
- BANKS, W. P. Signal detection theory and human memory. *Psychological Bulletin*, 1970, 74, 81-89.
- BATTIG, W. F. Facilitation and interference. In E. A. Bilodeau (Ed.), *Acquisition of skill*. New York: Academic Press, 1966.
- ELLIS, H. C., & DANIEL, T. C. Verbal processes in long-term stimulus recognition memory. *Journal of Experimental Psychology*, 1971, 90, 18-26.
- GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, 1940, 47, 196-229.
- GOTZ, A., & JACOBY, L. L. Encoding and retrieval processes in long-term retention. *Journal of Experimental Psychology*, 1974, 102, 291-297.
- GREENBERG, R., & UNDERWOOD, B. J. Retention as a function of stage of practice. *Journal of Experimental Psychology*, 1950, 40, 452-457.
- GREENO, J. G., JAMES, C. T., & DA POLITO, F. J. A cognitive interpretation of negative transfer and forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 331-345.
- IRION, A. L. The relation of set to retention. *Psychological Review*, 1948, 55, 336-341.
- IRION, A. L. Retention and warming-up effects in paired associate learning. *Journal of Experimental Psychology*, 1949, 39, 669-675.
- JOINSON, P. A., & RUNQUIST, W. N. Effects of intralist stimulus similarity and degree of learning on forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 554-559.
- MARTIN, E., FLEMING, F. G., & NALLY, P. D. Effects of temporal locus of a recitation attempt on learning and retention. *Memory & Cognition*, 1978, 6, 274-282.
- NELSON, D. L., BROOKS, D. H., & WHEELER, J. W. Sensory and meaning features in stimulus recognition and associative retrieval. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 711-719.
- POSTMAN, L. Transfer, interference, and forgetting. In J. W. Kling & L. A. Riggs (Eds.), *Experimental psychology* (3rd ed.). New York: Holt-Rinehart-Winston, 1971.
- RICHARDSON, J., & UNDERWOOD, B. J. Comparing retention of verbal lists after different rates of acquisition. *Journal of General Psychology*, 1957, 56, 187-192.
- RUNQUIST, W. N. Intralist interference and similarity. In C. P. Duncan, L. Sechrest, & A. W. Melton (Eds.), *Human memory: Festschrift in honor of Benton J. Underwood*. New York: Appleton-Century-Crofts, 1972.
- RUNQUIST, W. N. Conditional recall and bias in memory experiments. *American Journal of Psychology*, 1973, 86, 351-360.
- RUNQUIST, W. N. The assessment of discriminative encoding in paired associate learning. *Memory & Cognition*, 1974, 2, 472-478.
- RUNQUIST, W. N. Interference among memory traces. *Memory & Cognition*, 1975, 3, 143-159. (a)
- RUNQUIST, W. N. Stimulus structure, discrimination, and interference. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 491-500. (b)
- RUNQUIST, W. N. Physical similarity and stimulus recognition: Interference could still be associative. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 32-36.
- RUNQUIST, W. N., & SEKULICH, C. Transfer of stimulus differentiation and perceptual selection rules. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, 5, 326-336.
- RUNQUIST, W. N., & SNYDER, R. A. Forgetting of verbal paired associates after low degrees of learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 435-447.
- THOMPSON, C. P., WENGER, S. K., & BARTLING, C. A. How recall facilitates subsequent recall. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 210-221.
- TULVING, E., & PEARLSTONE, Z. Availability vs. accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 381-391.
- TULVING, E., & WATKINS, M. J. On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 181-193.
- UNDERWOOD, B. J. An evaluation of the Gibson theory of verbal learning. In C. N. Cofer (Ed.), *Verbal learning and Verbal Behavior*. New York: McGraw-Hill, 1961.
- UNDERWOOD, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 112-129.
- UNDERWOOD, B. J. The role of association in recognition memory. *Journal of Experimental Psychology*, 1974, 102, 917-939.

(Received for publication June 11, 1980;
revision accepted October 1, 1980.)