Interference in pigeons' long-term memory viewed as a retrieval problem

DAVID R. THOMAS, ALAN R. McKELVIE, MICHAEL RANNEY,
and THOMAS B. MOYE
University of Colorado, Boulder, Colorado 80309

In Experiment 1, three (experimental) groups of pigeons (n=8) acquired a successive wavelength discrimination in Phase 1 and a reversal in Phase 2; then, after a 24-h delay, they received a wavelength generalization test in extinction (Phase 3). For one group ("Context Same"), the same context was present throughout; for both others, a different context was used for Phase 1 and Phase 2. One group received the generalization test in the presence of Context 1, the other in Context 2. The Context Same and Context 2 experimental groups showed "recency," with all gradients peaking at the reversal S+ value. The Context 1 group yielded several different response patterns but never showed recency, thus revealing context-generated proactive interference. In Experiment 2, eight subjects learned the original discrimination and its reversal in different contexts, and each bird was tested alternately (within a session) in both contexts. Under this condition, the test contexts were effective retrieval cues. In every case, the gradients obtained in each context peaked sharply at the appropriate S+ value. These experiments indicate that conflicting memories may be stored along with their associated contexts such that they can be retrieved by an appropriate manipulation of contextual cues at the time of retention testing.

A generalization test may be viewed as a test of recognition memory in which responding maximally to a given stimulus value is analogous to a human adult subject making the verbal response, "That stimulus is the one to which I was trained to respond." Generalization gradients may be particularly useful measures of retention in situations in which there are potentially conflicting memories present. The typical procedure reported in the previous literature (cf. Spear, 1978) has been to train a rat on a discrimination problem (e.g., to turn right in a T-maze) and then to train the rat on its reversal (to turn left). An immediate test reveals a strong recency effect (i.e., the rat turns left). In a delayed test, interference from the first discrimination is revealed (i.e., the rat turns left less consistently and with longer latency). A problem with this procedure is that, since turning right and turning left are mutually exclusive responses, it is impossible to determine whether the observed interference is due to the strengthening of the right-turn tendency, the weakening of the left-turn tendency, or a combination of the two. A generalization test measure can resolve this question. If a pigeon is given successive discrimination training to peck at green but not at red, and then is trained on the reversal of this problem, the strength of the two response tendencies can be (relatively independently) assessed. Presumably, as a consequence of recency, a generalization test administered immediately after training would reveal a peak at red. If the later response tendency were to weaken relative to the initial response tendency during a delay interval, this could result, during a delayed generalization test, in the obtaining of a gradient with a (second) peak at green. This effect would be in addition to the nonspecific effect of forgetting on generalization gradients, that is, the flattening of the gradient with the passage of time between training and testing (cf. Thomas & Lopez, 1962) or with the inclusion of a purported source of interference (cf. Burr & Thomas, 1972).

In the present experiments, our intent was to use generalization gradients to make inferences about proactive interference (PI) in long-term memory in pigeons. These experiments are extensions of an earlier study reported by Burr & Thomas (1972). That study did not take advantage of the opportunity to separately assess the memories of original and reversal habits because the discriminative stimuli used were too closely spaced, but the experiment succeeded in demonstrating a strong PI effect. The

Burr and Thomas (1972) experiment involved eight different groups, one of which is critical for present purposes. In this group, the birds were first given successive discrimination training between yellow (VI reinforced) and green (extinguished) keylights for as many sessions as needed to achieve a rigorous criterion. Then they were trained on the reversal of this problem (green S+/yellow S−) during a single session that lasted until they mastered the reversal problem. Finally, they were tested for wavelength generalization in extinction after a 24-h delay. Proactive interference was reflected in the fact that this group yielded an extremely flat generalization gradient, reliably flatter than that of subjects given only the green S+/yellow S− problem or given both problems with this (second) problem learned under distributed practice conditions. The deficit was in long-term retention, since subjects tested only 1 min following these different training conditions produced gradients that were indistinguishable from one another.

The present experiments resembled that of Burr and Thomas (1972) in that they used a successive wavelength discrimination, trained under distributed practice, followed by its reversal, trained within a single session. The S+ and S− wavelength values were more widely spaced, however. More importantly, in the present experiments contextual stimuli were manipulated in an attempt to modulate the degree of PI obtained. The rationale for the study follows from Spear's (1971, 1978) theoretical account of animal memory, which attributes much of forgetting to the lack of appropriate or a sufficient number of retrieval cues at the time of retention testing. According to this position, ambient (contextual) stimuli, present but inconsequential to the target learning task, can function as "reminders," eliciting retrieval of a memory of prior training if they are stored as attributes of that memory. The recency effect—that is, the tendency of the subject to respond in accordance with most recent contingencies—is interpretable in terms of internal contextual stimuli (e.g., deprivation conditions, hormonal states, etc.) serving as effective retrieval cues. During retention testing, these stimuli would be more similar to what they had been in the immediately preceding training than to what they had been at a point further removed in time. In the remainder of this manuscript, we will use the term "recency cue" to refer to the unspecified stimuli that give rise to the recency effect.

Our past research (cf. Burr & Thomas, 1972) indicated the presence of a strong recency effect in an immediate generalization test following reversal training. That is, during a generalization test, birds tended to respond in a manner consistent with the most recent problem experienced. Interference was inferred from the degree to which subjects failed to respond in accordance with the recency principle.

Suppose that an original problem and its reversal were learned in different environmental contexts. Then, testing for generalization in Context 2 should minimize any interference from the original problem, since both the recency and the context cues are in accord. When testing is done in Context 1, however, interference should be maximal, because the recency and context cues are in conflict. Experiment 1 was designed to test these hypotheses, and it included control groups to determine the nature of the generalization gradients that would result from experience with only one of the two training problems, plus another control group that experienced both problems in the same context.

EXPERIMENT 1

Method

Subjects. The subjects were 40 experimentally naive pigeons obtained from a local supplier and maintained at 75% of their ad-lib weights. They were housed in individual cages in a colony room with a 16-h/8-h light/dark cycle.

Apparatus. The experimental chambers were two similar sound-attenuating ice chests bisected by aluminum "intelligence panels." The subjects' environment was 28.6 cm long × 28.6 cm wide × 24.1 cm high and had a 1.3-cm² hardware cloth floor. Access to the solenoid-operated food hopper was through a 3.8 cm high × 5.1 cm wide aperture centered on the intelligence panel 7.6 cm above the floor. The 1.9-cm-diam key was located 11.4 cm above the hopper and required .12 N of force to operate. Key stimuli were produced by IEE projectors (Model 00010-01-3043-1815) with No.1815 lamps operated at 12.5 V. Kodak Wratten filters Nos. 75, 74, 99, 73, and 72B produced nominal peak-transmitted wavelengths of 490, 538, 555, 576, and 606 nm, respectively. The houselight was a No. 1829 bulb operated at 24 V behind a 3.8-cm-diam light-diffusing Plexiglas disk in the extreme upper right-hand corner of the intelligence panel. A standard electronic tone generator provided white noise or a 1,000-Hz tone at 80 dB SPL (re: .0002 dynes/cm²). Exhaust fans provided masking noise. Reinforcement was 3 sec of access to the food hopper filled with mixed grain. All recording and control were accomplished with standard relay circuitry in an adjoining room.

Procedure. In preliminary training, subjects in the experimental groups were treated identically except for the ambient stimuli present. For 12 subjects, the context consisted of houselights-on plus a 1,000-Hz tone (HLT); for the other 12 subjects, the context was houselights-off plus white noise (H̅L̅N̅). All subjects were magazine trained, hand shaped to peck the 538-nm keylight, and allowed to earn 60 reinforcers on a gradually increasing variable interval (VI) schedule. On Day 2, subjects were given single stimulus training with the 538-nm keylight in their appropriate contexts until 40 reinforcers were earned on a VI 30-sec schedule. Starting on Day 3, discrimination training was given during which responding was reinforced on a VI 30-sec schedule during S+ trials and extinguished during S− trials. Subjects were run daily in 30-min sessions composed of three blocks of five S+ (538 nm) and five S− (576 nm) trials. Trials were of 57-sec duration, each separated by 3-sec blackouts, in a standard Gellerman series. Subjects were run on this task until 90% of their total responses were made in the presence of the S+ during a single session. When each subject reached criterion, it was shifted to Phase 2 discrimination training on the following day.

In Phase 2, all subjects were run in a single session to a criterion of 90% of total responses to the S+ for a block of 10 trials. Trials and blackouts were of the same duration as in Phase 1.

However, the discrimination problem was the reversal of that learned in Phase 1 (i.e., S+ was 576 nm and S− was 538 nm). The 24 subjects were divided into three groups of eight each, with four subjects in each group trained initially in each of the two contexts, HLT and $\overline{HLN}$. For one group (Context Same), the context in Phase 2 was the same as that used in Phase 1; for Groups Context 1 and Context 2, the alternative context was used with the reversal problem in Phase 2.

Eight subjects served in each of two control groups. One of these groups experienced only Problem 1 (538 nm S+, 576 nm S−), whereas the other experienced only Problem 2 (576 nm S+, 538 nm S−). Half of each group was trained in the HLT context and half in the $\overline{HLN}$ context. The principal purpose of these groups was to determine the form of the gradient and, most particularly, the location of its peak, when the subjects had experienced a single discrimination problem. No attempt was made to match control subjects to experimental subjects on the basis of total training time. Both control groups received the same pretraining as did the comparable experimental groups. The Problem 2 control group experienced distributed practice (i.e., daily 30-min sessions) until the 90% discrimination criterion was reached. The Problem 1 control group served the additional purposes of determining whether exposure to two different contexts and to both distributed and massed practice would have any marked effect on generalization. On the day after the Problem 1 control subjects achieved their initial discrimination criterion, they received an additional session with the same problem (538 nm S+, 576 nm S−) in the alternate context until the criterion of 90% of total responses to S+ within a 10-trial block was achieved. For these subjects, generalization testing was carried out in the context in which massed practice had been administered.

On the day after the completion of their training, the subjects in all groups were tested for wavelength generalization in extinction. The five test stimuli 490, 538, 555, 576, and 606 nm were randomized into blocks, and 12 different randomized blocks were presented to each subject. Stimulus presentations were of 30-sec duration separated by 5-sec blackouts.

Results and Discussion

There was considerable between-subjects variability in the time required to meet the discrimination criterion on Problem 1, ranging from two to five sessions. Similarly, time to criterion on Problem 2 had a substantial range (i.e., from 50 to 130 min). There was no systematic difference, within either Problem 1 or Problem 2, in time criterion as a consequence of whether the context employed was HLT or $\overline{HLN}$. There was, however, a difference in time to acquire the reversal, depending upon whether or not the context was changed between the two training phases. In the same context, a mean of 101.2 min was required to master the reversal, compared with 73.1 min in the altered context [t(22) = 2.16, p < .05]. This finding is consistent with the view that Context 1 was associated with the memory of Problem 1; thus, when reversal training was carried out in this context, the reactivated memory of Problem 1 interfered with the acquisition of Problem 2. Chiszar and Spear (1969) reported a similar observation in a rat study in which a spatial reversal was carried out in the same T-maze or in a different one.

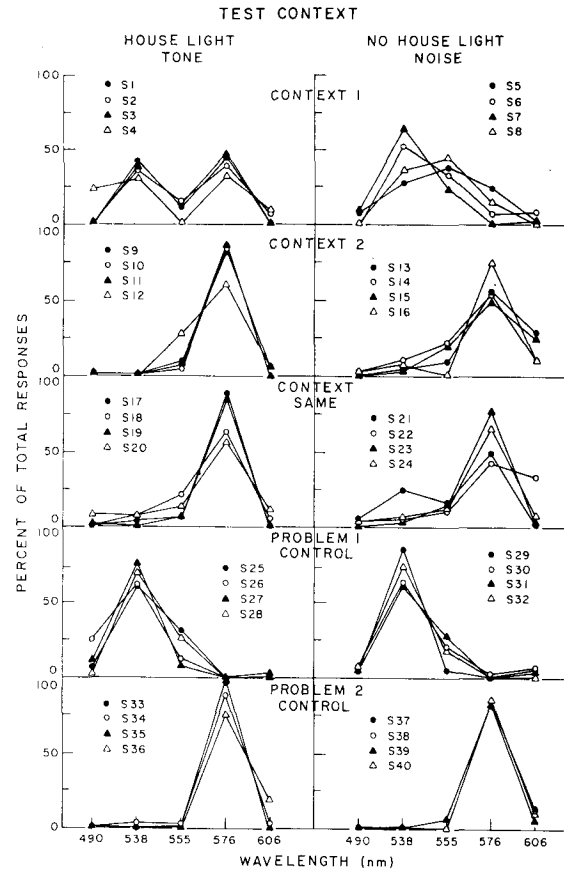Of primary interest are the results of the generalization tests, which are presented in Figure 1. The

Figure 1. Generalization gradients obtained following reversal or single problem training, in one or two different contexts, with testing carried out in the original context (Context 1) or the reversal context (Context 2).

two columns of the figure correspond to the two context conditions present during generalization testing, HLT in the left column and $\overline{HLN}$ in the right column. This variable was completely counterbalanced, so that half of the subjects in each treatment group were tested in each of these contexts. While Figure 1 presents the relative gradients, this is an accurate reflection of the absolute gradients also. In fact, if the ordinates of the mean relative gradients were relabeled 0 to 700 instead of 0 to 100, the relative and absolute plots would be virtually indistinguishable.

Consider, first, the gradients of the two control groups. The Problem 2 control group, presented in the bottom two panels of Figure 1, shows striking consistency. Regardless of the particular context experienced, all of these subjects yielded generalization gradients that peaked sharply at the training S+ value, 576 nm. The gradients of the Problem 1 control group, presented in the second two panels from the bottom in Figure 1, are also quite consistent. They are flatter than the gradients of the Problem 2 control group, but it is indeterminate whether this

is due to differences in discriminability within the wavelength continuum or to the exposure of these subjects to both contexts and to both massed and distributed training conditions. For present purposes, it is sufficient to note that all of the gradients in this group peak unequivocally at 538 nm, the appropriate S+.

Next, consider the gradients obtained from the Context Same group, shown in the center two panels of Figure 1. These subjects experienced both discrimination problems, yet their generalization gradients show a strong effect of recency; they are therefore somewhat similar to those of the Problem 2 control group. There is an indication of proactive interference from the memory of Problem 1 in these gradients, however. One subject (S21) yielded a gradient with a secondary peak at the S+ from Problem 1, 538 nm, and the gradients of this group tended to be flatter than those of the Problem 2 control group. A measure of the slope of the gradient is the percent of total responses emitted to the S+ value. The Problem 2 control group made a mean of 87.4% of their responses to 576 nm, compared with 66.6% in the Context Same group [t(14) = 3.20, p < .01]. Thus, experience with the prior reversal problem, in the same context, produced a significant PI effect.

It was expected that the Context 2 group would show less PI than would the Context Same group; although the difference in percent of responding to 576 nm was in the appropriate direction, with 68.6% emitted to 576 nm in the Context 2 group, the gradients of these two groups were, for practical purposes, indistinguishable. This might suggest that the context manipulation had no effect; however, the results of the generalization tests performed under Context 1, presented in the top two panels of Figure 1, argue convincingly otherwise. These gradients are difficult to characterize since several different patterns of responding emerged. One thing is clear from Figure 1, however: testing in the presence of the context of Problem 1 completely overcame the recency effect. Not one subject in this condition yielded a gradient that peaked exclusively at 576 nm, the Problem 2 S+. Four of the gradients were bimodal, with essentially equal responding to both former S+ values throughout the course of the test. Two other gradients peaked at an intermediate value (555 nm) that had never been reinforced (or even experienced) in training, and two peaked at 538 nm, the Problem 1 S+.

Before further pursuing the theoretical significance of the results obtained with the Context 1 group, we decided to replicate this treatment with all combinations of the contextual cues. Eight naive subjects were trained in the same procedure as the Context 1 experimental group. Four subjects were a direct replication, while the other four were trained with the alternate combinations of the contextual

stimuli (i.e., $\overline{\text{HLT}}$ and HLN rather than HLT and $\overline{\text{HLN}}$). All three types of gradient emerged again in this replication study, and, again, not one subject showed a recency effect.

The following conclusions may safely be drawn on the basis of the results of Experiment 1 and the replication group. When pigeons learn a discrimination problem, then learn its reversal, and then are tested for generalization after a 24-h delay, their gradients will normally show recency, peaking at the Problem 2 S+. Proactive interference may be revealed in a flattened gradient, relative to that of a control group that experiences only Problem 2. In the present study, a group of subjects that experienced the original and reversal problems in different contexts and was tested for generalization in Context 1 showed massive interference between the memories of the two problems. Under this condition, only six (of 16) subjects produced gradients similar to those obtained in a control group that had experience with only Problem 1. This indicates that the context present during generalization testing may serve to retrieve the memory of the problem learned in that context, but that generally the recency effect predominates.

We have no explanation of why HLT as Context 1 tends to produce double-peaked gradients, whereas no other condition we have tried does so. Similarly, $\overline{\text{HLN}}$ is clearly the most effective retrieval cue for the memory of Problem 1, producing appropriate peaks in four of six cases, but there is no apparent reason why other cues failed to work as well. As will be seen, rather than searching for more effective contextual cues in Experiment 2, we adopted the alternative strategy of performing an experimental manipulation in an attempt to make the cues we had already employed more salient.

The literature on discriminative stimulus control indicates that the effects of intradimensional successive discrimination training on stimulus generalization—in particular, on the sharpening of the gradient and the peak shift—require the alternation, within a session, of S+ and S− stimulus values (cf. Ellis, 1970; Honig, Thomas, & Guttman, 1959). This manipulation, at the time of training, presumably serves to call attention to the discriminative stimuli. We reasoned that a similar manipulation (i.e., the alternation of contextual stimuli *during generalization testing*) might increase the salience of *those* cues thereby revealing what had been learned about them in prior training. This supposition proved to be correct.

EXPERIMENT 2

Method

Subjects. The subjects were eight experimentally naive pigeons, maintained as in Experiment 1.

Apparatus. The same apparatus was used as in Experiment 1.

Procedure. The training procedure and discrimination problems were the same as those used for the Context 1 group in Experiment 1 for four subjects and those used in the replication group with four others. Thus, for four subjects Context 1 was HLT and Context 2 was $\overline{\text{HLN}}$. For the other four subjects, Context 1 was HLN and Context 2 was $\overline{\text{HLT}}$. One subject in the HLT group failed to respond during generalization testing and was replaced.

Generalization testing was carried out 24 h after the completion of training, and, as in Experiment 1, the five test wavelengths were randomized into blocks, and 12 different random blocks were presented to each subject. The context was changed, however, after each pair of blocks. This was done in a counterbalanced fashion such that for half of the subjects, Blocks 1 and 2, 5 and 6, and so on, were experienced in Context 1, whereas for the other subjects they were experienced in Context 2. Durations of stimulus presentations and blackouts were the same as in Experiment 1.

Results and Discussion

As in Experiment 1, the acquisition functions did not vary systematically with the particular contextual cues used in the two training phases. The mean times to criterion were 2.5 sessions for Problem 1 and 62.5 min for Problem 2. These times were comparable to those required in the original experiment and the replication. The generalization gradients of the eight subjects in Experiment 2 are presented in Figure 2. For the subjects in the top row, HLT was Context 1 and $\overline{\text{HLN}}$ was Context 2. For the subjects in the bottom row, HLN was Context 1 and $\overline{\text{HLT}}$ was Context 2. For all subjects, the generalization gradient obtained under Context 1 is plotted with closed circles, whereas the gradient obtained under Context 2 is plotted with open circles. As may be

seen in the figure, both contexts served as highly effective retrieval cues for all subjects. In every case, the gradient peaked at the S+ value appropriate to the context present at the time and showed essentially no responding to the S+ experienced in the alternative context. There was no systematic effect of whether the context first experienced during testing was Context 1 or Context 2. Similarly, it made little difference what the specific stimulus values were that constituted each context.

These results are what would have been expected if the birds had been extensively trained on an explicit conditional discrimination; for example, when houselight and tone are present, 538 nm is S+ and 576 nm is S−; when no houselight and white noise are present, the reverse is true. Note, however, that such conditional discriminations are difficult to establish in pigeons when traditional methods involving repeated alternations of the conditional (superordinate) stimuli are employed. In the present experiment, there was only one alternation *in training* between Context 1 and Context 2, and furthermore, the association between Context 2 and Problem 2 was formed in a single training session. Richards (1979) reported a study in which a conditional cue, a vertical black line, determined which of two colors was reinforced. In the presence of the line, white was S+, and red was S− (counterbalanced). In the absence of the line, the reverse was true. The four components of this conditional discrimination were successively alternated within sessions. Two of the original eight birds were dropped from the study for showing no learning after 30 sessions, and only
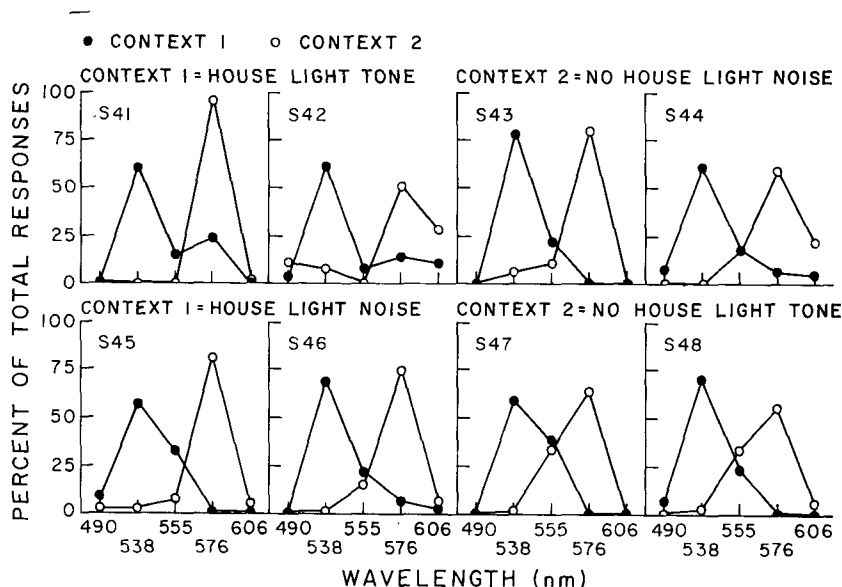

Figure 2. Generalization gradients obtained following original and reversal training in two different contexts, with the contexts alternated during the course of generalization testing.

two of the remaining six birds had achieved a discrimination ratio of greater than 90% of total responses to the S+ values after 75 sessions of training. Although no prior study has employed contextual stimuli like those used here as explicit conditional cues, it would seem unlikely that they would be more salient than are the on-key stimuli that have been used. This suggests that conditional stimulus control, by context or by on-key stimuli, is probably established far faster than previously realized, but a generalization test in extinction is required to reveal it. The greater sensitivity of generalization than discrimination performance measures of stimulus control has been known for some time. Thomas (1962) trained subjects on several different intradimensional wavelength discrimination problems and tested them repeatedly for generalization during the course of acquisition of the discrimination. The gradients revealed a shift in area away from the S− side of S+ prior to any measurable change in discrimination ratio.

GENERAL DISCUSSION

Given the magnitude of the PI effect obtained in the Context Same group of Experiment 1 and the clear evidence from Experiment 2 that both contexts had been associated with appropriate memories, it is unclear why testing in Context 2 failed to reduce PI in Experiment 1. It may be that the interference produced by providing these subjects with a potentially conflicting memory outweighed the benefits of having two different (though congruent) retrieval cues operating concurrently.

In Experiment 1, testing in Context 1 was sufficient to overcome the recency effect but not to retrieve very effectively the memory of Problem 1. In Experiment 2, experience with both contexts during generalization testing was effective not only in overcoming the effect of recency, but also in allowing the subjects to display appropriate retention performance in both contexts. Since the training procedures were identical in the two experiments, it must be concluded that the association between Context 1 and Problem 1 was formed equally strongly in both experiments, and the observed difference in performance at the time of retention testing must be attributed to differences in retrieval rather than in storage.

In conclusion, when training of two problems is done in a single context, that context, if present

during testing, would tend to retrieve both memories. If testing is done soon after training, the similarity of internal contextual cues in training and in testing may result in a recency effect. In delayed testing, the change in those internal cues may eliminate the recency cue, resulting in poor performance in a retention test. When different contexts are used in training, they are stored as attributes of the target memories such that later retention testing in each context may retrieve the memory associated with that context. These experimentally manipulated exteroceptive contextual retrieval cues may become more salient than the interoceptive ones that purportedly account for the recency effect. A remaining question, which may be unanswerable, is whether the use of different contexts is required for the separate storage of the memories of the two problems or, instead, serves exclusively to facilitate their selective retrieval. It is not clear whether (or how) we can distinguish between storage and retrieval functions of the exteroceptive contextual cues since they are so intimately interrelated.

REFERENCES

Burr, D. E. S., & Thomas, D. R. The effect of proactive inhibition upon the post-discrimination generalization gradient. *Journal of Comparative and Physiological Psychology*, 1972, 81, 441-448.

Chiszar, D. A., & Spear, N. E. Stimulus change, reversal learning and retention in the rat. *Journal of Comparative and Physiological Psychology*, 1969, 69, 190-195.

Ellis, W. R., III. Role of stimulus sequences in stimulus discrimination and stimulus generalization. *Journal of Experimental Psychology*, 1970, 83, 155-163.

Honig, W. K., Thomas, D. R., & Guttman, N. Differential effects of continuous extinction and discrimination training on the generalization gradient. *Journal of Experimental Psychology*, 1959, 58, 145-152.

Richards, R. W. Stimulus control following training on a conditional discrimination. *Animal Learning & Behavior*, 1979, 7, 309-312.

Spear, N. E. Forgetting as retrieval failure. In W. K. Honig & P. H. R. James (Eds.), *Animal memory*. New York: Academic Press, 1971.

Spear, N. E. *The processing of memories: Forgetting and retention*. Hillsdale, N.J: Erlbaum, 1978.

Thomas, D. R. The effects of drive and discrimination training on stimulus generalization. *Journal of Experimental Psychology*, 1962, 64, 24-28.

Thomas, D. R., & Lopez, L. J. The effect of delayed testing on generalization slope. *Journal of Comparative and Physiological Psychology*, 1962, 44, 541-544.