Interference in memory for multiple contexts*

JOHN R. ANDERSON† University of Michigan, Ann Arbor, Michigan 48104

and

GORDON H. BOWER Stanford University, Stanford, California 94305

Remembering that an item occurred in several different lists is formulated here in terms of retrieval of corresponding list tags associated to the item. Therefore, associative interference should operate upon remembering the several list contexts in which an item appeared. Experimental Ss studied four (or five) overlapping lists of 16 words, sampled from a master set of 32 words, with a given word exemplifying one of the 2⁴ (or 2⁵) possible sequences of appearances and nonappearances over the four (or five) lists. Later Ss rated from memory for each word and for each list whether that word had occurred in that list. After correcting for interlist generalization effects, indices of discriminative memory revealed strong proactive interference and weaker retroactive interference. Discriminative memory that an item occurred in a given list was poorer the more prior or more subsequent lists in which that item had also occurred. Thus, list differentiation appears explicable in terms of item-specific associative interference.

A recent trend in memory research concerns analysis of S's memory for the environmental "context" surrounding the occurrence of a discrete event. In the laboratory, the typical event studied is presentation of a single verbal item in a given context; on the later memory test, S is shown the item again and asked to remember something about the earlier context of its occurrence. Results have shown that Ss apparently store a relatively full "description" of the input event, even under incidental learning conditions. The performance almost resembles that of a video or audio tape recorder, albeit a somewhat faulty one. Examples of the types of contextual information that Ss can remember include whether the item was presented visually or auditorily, in uppercase or lowercase letters, in a male or female voice, as a word or picture, on the left or right side of the visual display, alone or in company of another item, for a short or long exposure time, and so forth (e.g., Hintzman, 1971; Light, Stansbury, Rubin, & Linde, 1973). Moreover, Ss can remember the approximate temporal position within a list at which an item occurred (Hintzman & Block, 1971) and, in multilist experiments, the list and the position within the list at which an item occurred (Hintzman, Block, & Summers, 1973).

The foregoing judgments reflect aspects of context retrievable following a single presentation of an item. It has also been demonstrated that Ss can recall the fact of multiple occurrences of a single item; they can judge fairly accurately the frequency of an item's occurrences

*This research was supported by Grant MH-13950 from the National Institutes of Mental Health to G. H. Bower. We wish to thank Bob Rothbart for assisting with Experiment II and Lynne Reder for her comments on this manuscript.

Reder for her comments on this manuscript. †Reprint requests should be sent to John R. Anderson, Human Performance Center, 330 Packard Road, Ann Arbor, Michigan 48104. in a list (e.g., Underwood, Zimmerman, & Freund, 1971), the several positions within a list at which a repeating item occurred (Hintzman & Block, 1971), the temporal spacing between repetitions of an item in a list (Hintzman & Block, 1973), and the frequencies of an item's presentations in each of two lists (Experiment III of Hintzman & Block, 1971).

A logical next step in trying to understand context memory would begin analysis of interactions between the several contextual memories retrievable from a multiply presented item. A recent formulation by Anderson and Bower (1972, 1974) suggests one approach to studying such interactions. They suppose that, upon presentation of a known verbal item, its internal representation (in long-term memory) becomes associated with a "context marker," which is conceived to be a collection of descriptors of the context surrounding that item's presentation. It is an empirical matter to discover the content of these descriptors and how they vary with the mental set of S; but they would include such descriptors as the item's physical appearance, the prevailing list name, internal associates set off by the item, and whatever internal elements permit "time tagging."

For present purposes, the content of these context descriptors is not so important as is the assumption that an association is formed from an internal representation of the item (call it Node A in a memory network) to another node which collects together and represents the context prevailing during this item's presentation (call this Node B). That is, presentation of item A in context B is represented in memory by formation of an A-B association much as in paired-associate learning. This association encodes the proposition "Item A occurred in

context B," with the understanding that context B itself has descriptors d_1, d_2, \ldots . This formulation suggests that interference principles relating to transfer and retroaction (see Keppel, 1968) may also apply to such context associations. Whenever a single item (A) occurs in successive contexts (say, lists L_1 , L_2 , L_3 of an experiment) so that associations A- L_1 , A- L_2 , A- L_3 are being successively established, we may ask whether there is negative transfer in learning as well as retroactive and proactive interference operating upon retention of particular context associations.

Such an application of interference theory to list discrimination must proceed with some caution, however, because the contexts may be similar, resulting in generalization of associations between contexts. Thus, if list contexts L₁ and L₂ are similar and item A occurs only in List 2, the association A-L2 may generalize, leading S to give a false positive response to the test question "Did word A occur in List 1?" In a homogeneous multilist experiment, such interlist generalization is known to be greater the closer two lists are in time (see, e.g., Hintzman et al, 1973). Moreover, the net tendency to give a false positive response to a given test association (say, A-L₂) should vary directly with the sum of generalized strengths of associations to that item from neighboring lists (say, A-L₁ and A-L₃). As an illustration, in a two-list experiment by Hintzman and Block (1971, Experiment III), a given item was presented zero, two, or five time in List 1 and, orthogonally, zero, two, or five times in List 2. Later judgments of "frequency in list" showed consistent interlist generalization; holding constant an item's actual frequency in (say) List 2, its List-2 frequency as judged from memory was higher the more frequently the item had occurred in List 1. There was a similar generalization effect of an item's List-2 frequency upon judgments of its List-1 frequency.

Now, such generalization effects complicate tests of interference principles in list discrimination simply because the two effects act in opposite directions upon responses indicative of context memory. To illustrate, in a three-list experiment in which item A occurs in all three lists, whereas item B occurs only in List 2, interference principles imply that the A-L₂ association should be learned less (and forgotten more later) than the B-L2 association, so on that basis alone there should be fewer "positive" responses to A than to B when the question is whether the item occurred in List 2. But, since item A also occurred in Lists 1 and 3, the associations A-L₁ and A-L₃ will generalize and summate, producing a high net strength to A-L₂ (see Bower, 1972, for a recent mathematical model of this process). These sources of generalized strength are not available for item B, which occurred only in List 2; so on this basis, one expects less positive responses to a test question regarding B-L2 than to one regarding A-L2.

A possible solution is to estimate the tendency to give generalized responses to a particular list that has been established by an item's pattern of occurrences in other lists and then use that estimate to "correct" the raw probability that S thinks he remembers the item occurring on that particular list, when in fact it did. Thus, S's memory for A-L2 when A has occurred in Lists 1, 2, 3 would be "corrected" by use of S's false positive rate to C-L2, where item C occurred in Lists 1 and 3 (as did A) but not in List 2. The actual "correction" used below is the standard d' of signal detection theory (see, e.g., Anderson & Bower, 1972), where the confidence ratings regarding the truth of the probe A-L2 are taken to correspond to the "signal distribution," whereas the ratings for C-L2 are taken as the appropriate "noise distribution." The d' so derived could be interpreted as the discriminability of an item's presentation (vs nonpresentation) in a given list, given a particular history or pattern of its occurrences in other lists preceding and/or following the specific list under consideration. These d' measures may be considered as measures of memory for list context corrected for interlist generalization; as such, they may be examined directly for evidence of interference effects.

EXPERIMENTS I AND II

In Experiment I Ss studied four overlapping lists of 16 words sampled from a master set of 32 words. Across the four overlapping lists, there are 24, or 16, possible sequences of appearances and nonappearances for a particular word. Two words from the master set of 32 were used to realize each of the 16 sequences, resulting in 16-item study lists. The test list was composed from all 32 words of the master set. Ss were asked to indicate for each test word and each study list whether that word had occurred in that list and to rate the confidence of their answer. Because of the list construction method, appearance or nonappearance of a particular item in any list was no predictor of its appearance in any other list. Experiment II was a simple generalization of this design, performed to determine the replicability of the results in Experiment I. Ss studied five lists of 16 words in which all 25, or 32, sequences of appearance and nonappearances of words were realized once. Except for the added list in Experiment II, the two experiments were identical in procedural details.

The question of interest is how memory for a word's occurrence in a particular target list is affected by its appearance in other lists. The interference analysis predicts that the item's appearance in prior lists makes discrimination harder with respect to the target list due to negative transfer and proactive interference. Similarly, the appearance of the word in lists following the target list should cause retroactive interference in retrieval of the appropriate list tag. Therefore, list discrimination for a particular target list should deteriorate as the word appears in more nontarget lists.

Method

Material. Three master sets of 32 common concrete nouns were created for the experiment. From each master set, five lists of 16 words were created, with one word assigned to realize each of the 32 possible sequences of appearance and nonappearance across the five lists. For Experiment I only the first four lists were employed from a set. The words were typed onto slides. The slides for a particular list were given a distinct background different from that for other lists. The five backgrounds chosen were patterns of stripes, of dots, of circles, of stars, and of bricks. Three sequences of five lists were created in this manner, one from each of the three master sets. The order in which the master sets were tested, order of lists within a master set, and order of words within a list were randomized across experimental sessions.

Procedure. The words in a list were presented at a 5-sec rate, with a 20-sec interval between lists. At the beginning of each list, the E said, "This is the beginning of List N" and at the end of the list he said, "That is the end of List N," where N was the ordinal number of the list. Immediately after the presentation of the last list, the test procedure began, projecting, one at a time, each word from the master set for a 30-sec period. While the test word was displayed, the S had to make four (or five in Experiment II) judgments regarding whether the word had occurred in each of the four (or five) lists. For each judgment he was asked to assign a confidence rating between 1 and 3. A confidence rating of 1 indicated the S was very uncertain, a rating of 2 indicated some confidence in the decision, and a rating of 3 indicated the S was completely certain. After having rated the 32 words, the S was immediately shifted to the next master set for study and testing. For later scoring, the confidence ratings for "yes" responses were assigned positive values 1, 2, or 3, and those for "no" responses negative values,

yielding a 6-point scale. Subjects. Thirty-six Ss were tested in Experiment I in three groups of sizes 10, 11, and 15. Thirty-eight Ss were tested in Experiment II in five groups ranging from 5 to 10. These Ss were recruited from the Stanford student body and were paid \$3.50 for their services in the experiment. The Ss in Experiment I were enrolled in the summer session at Stanford and those in Experiment II were enrolled in the following spring session.

Results and Discussion

For all analyses, the relevant question is how S's memory (and judgment) that an item occurred in a particular target list depends upon the pattern of its appearances and nonappearances in the other three (or four) nontarget lists. We will consistently use the terms target and pattern in this way. As indicated before, we must examine discriminability (d') between items having the same pattern except for their appearance or nonappearance on the target list. Before proceeding to those discriminability scores, let us first examine the evidence for interlist generalization in "raw" confidence ratings.

A first fact is that the S's average rating that an item occurred in a target list was higher the greater the number of nontarget lists in which the item had also occurred. Considering only items that in fact appeared in the target list and pooling over lists, in Experiment I the mean confidence rating for target appearance of an item that appeared in zero nontarget lists was .88, in one nontarget list .90, in two nontarget lists 1.06, and in three nontarget lists 1.52. A similar monotonic trend occurred in Experiment II, with average ratings for

target appearance of an item being .86, .88, 1.07, 1.12, and 1.41 for its appearance in zero, one, two, three, or four nontarget lists, respectively. "False positive" ratings for items that did not occur in a target list similarly increased with the number of nontarget lists in which that item had occurred.

Furthermore, there is somewhat of a "temporal gradient" in these interlist generalization effects. An estimate of generalization from, say, List 2 to 3 may be obtained by comparing the mean confidence rating for a List-3 appearance for all patterns which have a List-2 appearance to the similar mean List-3 rating for patterns which do not have a List-2 appearance. The difference between these two mean ratings measures the effect of the appearance of a word in List 2 upon S's confidence that it appeared in List 3. Such measures, obtained for all list pairs, are shown in Fig. 1, for Experiment I in Panel a and for Experiment II in Panel b. The abscissa is the list being rated and the parameter labeling each curve is the list from which generalization is being computed (i.e., for which effects of appearance vs nonappearance are being compared). The peaks of the curves reflect accurate discrimination. As can be seen, there are slight temporal generalization gradients; the effect falls off monotonically with distance. Although the decline in this index between distances of 1, 2, and 3 is small, it is very consistent throughout all the curves. Excluding the peak, there are six comparisons of means in Experiment I at distances of i and i + 1 (for i = 1, 2). All six of these comparisons are in the expected order. Experiment II provides 12 such adjacent comparisons, 11 of which are in the expected direction. There thus appears to be evidence for interlist generalization varying with the temporal distance between lists. This finding corroborates an earlier one by Hintzman et al (1973). Such temporal gradients may be interpreted in terms of continuous changes over time in the set of available contextual elements, with more overlap in sets the closer they are in time. Therefore, list markers for adjacent lists would be comprised of many similar elements and hence more confusable (see Bower, 1972, for a model of this process).

Signal Detectability Measures of Discrimination

As noted in the introduction, the existence of generalization effects indicates the need for a discriminability index like d' (see Green & Swets, 1966) of memory for a word's appearance in a particular target list given a particular pattern for its nontarget appearances. Sets of d' measures were obtained for each list using the confidence ratings. This was accomplished by comparing pairs of distributions of confidence ratings with respect to that target list, one pair for each possible pattern of nontarget lists. Ratings for words that occurred in a target list with a particular pattern were compared to ratings for words that had not occurred in the target but had otherwise occurred in the same pattern. For instance, in Experiment I, with List 3 as the

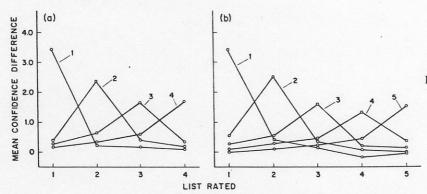


Fig. 1. List generalization gradients: (a) Experiment I and (b) Experiment II.

target, we would compare List 3 confidence ratings for words that had appeared in 1, not in 2, in 3, and in 4 with ratings for words that had appeared in 1, not in 2, not in 3, and in 4. The signal detectability analysis was performed with the first set of confidence ratings viewed as being produced from the signal distribution and with the second set corresponding to the noise distribution. Since both sets of words occurred in the same pattern of nontarget lists, list generalization from nontarget lists should be equated. Therefore, the d' measure of the mean distance between the signal and noise distributions should index the S's ability to remember and discriminate the occurrence from the nonoccurrence of a word in a target list when that word has occurred in a particular pattern of nontarget lists.

In Experiment I, considering each list in turn as the target, there were eight patterns of appearance and nonappearance of the word in the other three lists. Therefore, eight separate d's were calculated for each target list. With four target lists there were altogether 32 corresponding d's. In Experiment II 80 d's (five target lists times 16 patterns) were calculated. Moreover, these measures were computed separately for each master set. Thus, from Experiment I 96 d's were obtained from the three master sets, and from Experiment II 240. These measures are the data for the following statistical analyses.

One measure of list discrimination, d', is the intercept of the ROC curve with the negative diagonal when plotted on a normal-normal graph (see Clarke, Birdsdall, & Tanner, 1959). It is relatively insensitive to changes in the ratio of signal-to-noise variances. This is important, since in our data this ratio varied across conditions. We used the EPCROC program based on the method set forth by Ogilivie and Creelman (1968), which provides maximum likelihood estimates of signal detection parameters from rating data.

A first fact revealed by these analyses (and anticipated in the declining peaks of Fig. 1) is that a list's ordinal position strongly affected its discriminability. The mean d' estimates for Experiments I and II as a function of list position are displayed in Fig. 2. List discrimination clearly deteriorates across successive lists. Hintzman et al (1973) reported a similar "primary effect" in multilist

discrimination. This will be reflected in later estimates that proactive interference and negative transfer were more potent than retroactive interference in these experiments. That is, if a word appeared in a list preceding the target list, a greater decrement was caused in the word's target-list discrimination than if it had appeared in a subsequent list. However, even controlling for the number of preceding and subsequent lists which contained the critical word, discrimination was still worse for later lists. For instance, considering only words that appeared in two lists with their first appearance in List 2 or in List 3, the word which appeared for the first time in List 2 was more accurately identified as a List-2 word than the word which appeared first in List 3 was identified as a List-3 word (d' = 1.27 vs .94 for Experiment I and d' = 1.52 vs 1.31for Experiment II). One explanation of this primacy effect across lists would relate it to the relative distinctiveness of contextual cues near the beginning of the study tasks (an interpretation favored by Hintzman et al, 1973); another explanation would attribute the effect to some sort of generalized proactive interference that is not item specific. Both explanations seem plausible but sufficiently post hoc to warrant suspension of credibility.

Interference Effects in List Identification

Of crucial concern is the influence of a word's

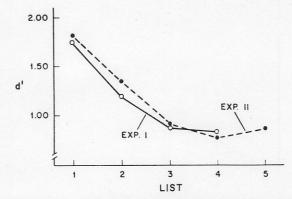


Fig. 2. List discrimination in Experiments I and II as a function of ordinal position of list.

appearance in other lists on the S's ability to discriminate its appearance in a target list. By design, these pattern-specific effects are orthogonal to the overall level of discriminability for particular lists. Tables 1 and 2 report the relevant data. The measure used is D, which is the algebraic difference of list discrimination (d') with respect to target List I for those words which occurred in nontarget List J minus the similar d' for those words which did not occur in List J. Since this is a complicated index, let us illustrate calculation of D for Experiment I when I is 1 and J is 2 (which yielded D = .399). First of all, for List 1 of Experiment I, 24 d's had been obtained (three master sets times eight patterns). We then subtracted the mean d' for the 12 patterns (three master sets times four patterns) in which the word did not occur in List 2 from the mean d' for the 12 patterns in which the word did occur. This yields an index of the average effect on a word's List-1 discriminability of its having also occurred in List 2. In Tables 1 and 2, arrayed down the rows are the target lists and across the columns are the nontarget lists (for which one is contrasting appearance vs nonappearance of a word). A positive value of D indicates that appearance in the nontarget list improved discrimination with respect to the target. The right-hand margin for each target list shows the average effect of the word's appearance in a nontarget list. The bottom row shows the average effect associated with each of the nontarget

The variability of the data across the three master sets was used to estimate the mean error. The standard error of the mean differences in Table 1 is .097; the same standard error applies to Table 2. This error estimate has 56 df for Experiment I and 150 df for Experiment II. Along with the D measures, Tables 1 and 2 also report t statistics evaluating the significance of these differences. Of course, many pairs of the t tests in these tables are not independent.

The results obtained in Experiments I and II are remarkably similar. In both experiments there is a highly significant decrement in target-list discrimination when

Table 1
Mean Effect on d' of Presence vs Absence of Target Word on Nontarget Lists: Experiment I (24)

Target List						
		1	2	. 3	4	Mean
1	D t		.399 4.11	097 1.00	106 1.09	.068 1.21
2	D t	016 .16		350 3.60	169 1.74	178 3.17
3	D t	121 1.25	323 3.32		245 2.53	230 4.10
4	D t	018 19	317 3.26	366 3.77		237 4.22
Mean	.D t	052 .93	080 1.43	271 4.83	173 3.08	144 5.13

the word also appears in a nontarget list (i.e., negative D values). While some of the t tests are not significant, the overall trend is highly significant. The average decremental effect of a word's appearance in a nontarget list is -.144 d' in Experiment I and -.121 d' in Experiment II. This decremental effect appears for all target lists except the first, where there is a nonsignificant trend in the opposite direction. Most nontarget lists cause decrements in list discrimination, but the effect is larger for lists that precede the target list (the Ds in cells below the main diagonal) than for lists that follow it (the Ds in cells above the diagonal). In Experiment I the mean effect of a preceding list was -.194 d' and of a subsequent list -.095 d'; in Experiment II the difference was even more marked. Subsequent lists actually have a slight positive effect (+.005 d'), whereas preceding lists have a large negative effect (-.246 d'). Thus, in these data proactive effects (negative transfer and proactive interference) are much more potent than retroactive effects (retroactive interverence).

A possible explanation for this asymmetry lies in a pattern of rehearsal that Ss commonly reported in this experiment. By the second or third replication through the multilist paradigm, the S knew that he had to

Table 2

Mean Effect on d' of Presence vs Absence of Target Word on Nontarget Lists: Experiment II (2⁵)

Target List			Nontarget List						
		1	2	3	4	5	Mean		
1	D		.047	.281	.024	142	.053		
	t		.48	2.87	.025	1.45	1.08		
2	D	361		.000	.053	086	099		
	t	3.69		.00	.54	.88	2.02		
3	D	358	396		010	104	217		
	t	3.66	4.05		.10	1.07	4.43		
4	D	451	118	267		018	214		
	t.	4.61	1.21	2.73		.18	4.38		
5	D	130	191	141	044		127		
	t	1.33	1.95	1.44	.45		2.60		
Mean	D	325	165	032	.006	088	121		
	t	6.64	3.37	.65	.12	1.80	5.53		

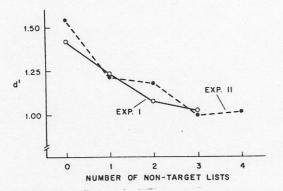


Fig. 3. List discrimination for words as a function of the number of non-target lists in which they have appeared.

remember the "list tags" for each word. Therefore, when a word appeared in List N, S would implicitly review and rehearse the names of earlier lists in which that item had also occurred. While such a rehearsal strategy enhances later remembering of early list tags, it does so at the expense of poorer learning of later list tags for the item. While this hypothesis explains the large "primacy" effect (alternatively, the larger proactive than retroactive effects), it remains post hoc until further experiments try explicitly by instructions or payoffs to manipulate rehearsal strategies.

As a final analysis of interference effects, Fig. 3 summarizes the effects on list discriminability of the number of nontarget lists in which the word had appeared. The data were pooled over all target lists and patterns of nontarget lists. Figure 3 reveals a strong decline in list discrimination with increasing numbers of interfering lists. The effects are almost identical for the two experiments. Of course, this is as expected by an associative interference analysis of multilist discrimination.

Let us summarize the significance of these experiments. As noted, Ss can remember fairly well the differing list contexts in which an item appeared across a multilist experiment. If we represent each such context memory in terms of a word-to-list tag association, then principles of item-specific negative transfer and associative interference should apply to the remembering of context information. Discriminability measures are required to gauge interference effects because interlist

generalization produces opposing effects on the simple recognition response. Using such list discrimination indices, both of our experiments demonstrated marked proactive and slight retroactive interference which was item specific. To our knowledge, this is the first clear demonstration of such item-specific interference effects in list discrimination. The success of the associative interference analysis in explaining interaction among memories for multiple contexts of an item indicates that "list differentiation" no longer needs to be viewed as an independent theoretical construct in explanations of retention in interference experiments (see, e.g., Postman, 1961). Happily, "list differentiation" itself turns out to be explicable in terms of associative interference principles.

REFERENCES

Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 79, 97-123.

Anderson, J. R., & Bower, G. H. A propositional theory of recognition memory. Memory & Cognition, 1974, 2, 406-412.
Bower, G. H. Stimulus-sampling theory of encoding variability. In A. W. Melton and E. Martin (Eds.), Coding theory in learning and memory. Washington, D.C: Winston, 1972.
Clarke, F. R., Birdsall, R. C., & Tanner, W. P., Jr. Two types of ROC curves and definitions of parameters. Journal of the Acoustical Society of America, 1959, 31, 629-630.
Green, D. M., & Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966. Anderson, J. R., & Bower, G. H. A propositional theory of

Green, D. M., & Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.
Hintzman, D. L. Effects of repetition and exposure duration on memory. Journal of Experimental Psychology, 1970, 83,

Hintzman, D. L., & Block, R. A. Repetition and memory: Evidence for a multiple trace hypothesis. Journal of Evidence for a multiple trace hypothesis. Journal of Experimental Psychology, 1971, 88, 297-306.

Hintzman, D. L., & Block, R. A. Memory for the spacing of repetitions. Journal of Experimental Psychology, 1973, 99, 70.74

70-74.

Hintzman, D. L., Block, R. A., & Summers, J. J. Contextual associations and memory for serial position. Journal of Experimental Psychology, 1973, 97, 220-229.

Keppel, G. Retroactive and proactive interference. In T. R. Dixon and D. L. Horton (Eds.), Verbal behavior and general behavior theory. Englewood Cliffs, N.J.: Prentice-Hall, 1968. Light, L. L., Stansbury, C., Rubin, C., & Linde, S. Memory for modality of presentation: Within-modality discrimination. modality Memory & Cognition, 1973, 1, 395-400.

gilivie, J. C., & Creelman, C. D. Maximum-likelihood estimation of receiver operating characteristic curve parameters. Journal of Mathematical Psychology, 1968, 5,

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), Verbal learning and verbal behavior. New York: McGraw-Hill, 1961.

Underwood, B. J., Zimmerman, J., & Freund, J. S. Retention of frequency information with observations on recognition and recall. Journal of Experimental Psychology, 1971, 87, 149-162.

> (Received for publication Oct er 18, 1973; revision received Januar, 1, 1974.)