AMERICAN JOURNAL OF PSYCHOLOGY
March 1976, Vol. 89, No. 1, pp. 127–135

A nonmonotonic lag function for false alarms to associates

Colin M. MacLeod and Thomas O. Nelson
University of Washington

In a continuous-recognition paradigm, undergraduates were asked whether they had already heard in the series each of the 300 words. The rate of false alarms to semantic associates of prior words depended strongly on the lag between the word and its associate: at no lag, there was absolutely no increase over the general rate; at a greater lag, the false alarms to associates increased; at still greater lags, they returned toward the general rate. This nonmonotonic lag function disconfirms all one-step notions of generalization, including those based on implicit associative responses or on feature overlap, whether activated during study or test, and it suggests several two-step models of forgetting during the lag.

Since the study by Underwood (1965), Shepard and Teghtsoonian's (1961) continuous-recognition paradigm has frequently been used to investigate the role of semantic relatedness in recognition memory (Anisfeld and Knapp, 1968; Fillenbaum, 1969; Grossman and Eagle, 1970; Kimble, 1968; Vogt and Kimble, 1973; Zimmerman and Kimble, 1973). Interest has centered on the kind of 'semantic generalization' needed to explain the increased rate of false recognitions of (false alarms to) those test words that are semantically related to prior words in the series. How are we to account for the finding that the subject's likelihood of saying 'old' to CHAIR, for example, increases if one of the prior words in the series was TABLE?

Nearly all of the current explanations postulate a *one-step* process, which varies in form from explanation to explanation. For some (e.g., Underwood, 1965), generalization occurs during study via implicit associative responses: that is, as the subject encodes TABLE, he implicitly activates the associate CHAIR. For others (e.g., Anderson and Bower, 1972), the implicit associative response occurs during test: that is, the subject implicitly activates the old word TABLE while he is being tested on CHAIR. This last explanation, however, is weakened by the evidence on the direction of association at study versus test (see Anisfeld and Knapp, 1968, Experiment II). For still others (e.g., Anisfeld and Knapp, 1968; Bower, 1972), generalization occurs during study via activation of features or elements the various words have in common: that is, semantic features such as 'type of furniture' might be activated (tagged) in CHAIR when the subject encodes TABLE.

All of these explanations implicitly or explicitly assume that memory strength (familiarity) is mapped into some kind of decision process based on a Thurstone scale (e.g., Egan's 1958 application of the theory of signal detection) and that the rate of false alarms to items that are semantically related to prior items is directly related to the memory strength of those prior items. This is made evident by Figure 1, in which the right side shows one possible representation of the memory process — as generalization during study via activation of features or elements — and the left side maps this representation into the decision process for responding in a continuous-recognition paradigm. Notice at the top right that immediately after study, five elements are activated in an old item (O), two elements are activated in a new related item (N_R; elements activated by the new item's similarity to the old item), and no elements are activated in a new unrelated item (N_U). These three kinds of items are entered into unidimensional distributions, shown at the top left of Figure 1, and a decision criterion determines whether the subject will respond 'old' or 'new.' Further, although only the elements-activated-at-study notion is shown, both of the other possibilities mentioned above would yield the

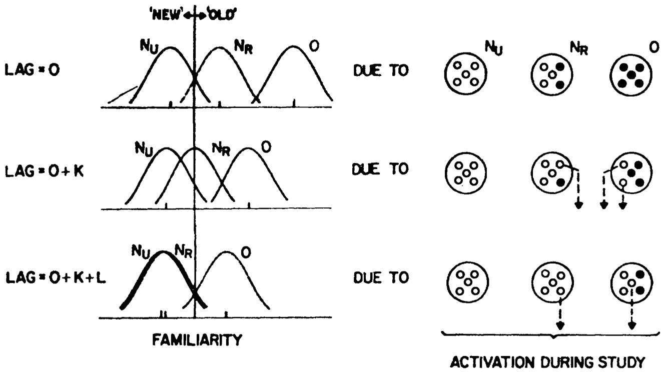

Figure 1. Comparison of three theoretical distributions of familiarity of the new unrelated (N_U), the new related (N_R), and the old (O) items at various study/test lags (0, 0 + K, 0 + K + L), as based on the number of elements activated at the time of study

same patterns of distributions shown in Figure 1; the same conclusions would also follow if recognition was based on other kinds of decision processes (e.g., Luce, 1963).

As additional items are presented, the numbers of elements shown activated at the middle and lower right of Figure 1 decrease (the dashed arrows), decreases accompanied by corresponding shifts in the distributions for the old and for the new related items. (The criterion is assumed constant, as in a block of trials where various lags are intermingled; see Underwood, 1965.) These shifts will produce successively fewer 'old' responses both for the old and for the new related items as the lag increases between a given old item's study and that old item's test (or new related item's test). Thus, as lag increases, there should be a *monotonic* decrease both in hits to old items and in false alarms to new related items, the latter eventually asymptoting at the general rate of false alarms to new unrelated items. Numerous previous studies have confirmed such a prediction of a monotonic decrease in hits to old items as lag increases (e.g., Raser, 1972). However, of all the studies examining recognition of semantically related words, none has systematically varied lag to test the straightforward prediction of a monotonic decrease in false alarms to new words semantically related to prior words. The present experiment, using semantic associates, tested this prediction.

METHOD

Subjects

The subjects were 50 University of Washington undergraduates who received bonus points in introductory psychology for their participation. The experiment was conducted using small groups of from two to seven students, determined by their availability.

Materials and apparatus

The associates used were the 25 highest forward-associate pairs in Palermo and Jenkins' norms (1964; e.g., LION-TIGER, TABLE-CHAIR), with the restriction that the associations not be synonyms, antonyms, superordinates, or subordinates; these were the 'associate pairs.' All of the remaining items, both those to be repeated and those to be presented only once, were chosen from the A and AA words in Paivio, Yuille, and Madigan's norms (1968), disregarding imagery and meaningfulness values. All items were nouns from three to seven letters long.

Of the 80 items to be repeated, 25 were selected to be treated in exactly the same manner as were the 25 associate pairs; these were the 'repetition pairs.' In the case of the associates, five pairs were presented at each of four lags — 0, 5, 10, and 15 intervening items — and the remaining five were control, in which only the second member of the associate pair was in fact presented. In the case of the 25 selected repetitions, five were presented at each of the four lags; the remaining five were control, in which the item was in fact presented only once. These associate and repetition controls were included in order to estimate the general rate of false alarms to these particular items.

Five experimental lists were composed; each list contained 300 words and was constructed in the following way. The first 15 and last 10 items were always once-presented items that acted as list buffers; within the list, there were 75 other once-presented filler items (a total of 100 filler words). Also within the list were 55 items repeated at various nonsystematic lags (a total of 110 words). The remaining 90 words were made up of the 20 associate pairs along with their 5 controls (45 words) and the 20 items repeated at systematic lags along with their 5 controls (45 words).

The 25 associate pairs and the 25 repetition pairs were each initially divided at random into five blocks of five items per block for purposes of counterbalancing and to avoid confounding with a possibly increasing rate of false alarms across the session (see Underwood, 1965). A 5×5 greco-latin square was used to ensure that each block appeared only once as block i ($i = 1$ to 5) across the five lists and that the items within each block occurred only once at position j ($j = 1$ to 5) within that block. Also, a 5×5 latin square was used to assign lags to the items within a block such that each item appeared once at each of the four lags and once as a control over lists. The same counterbalancing procedure was applied to both the associate and the repetition pairs; since this produced the same five sequences for repetitions and associates, a further randomization was done to prevent identical repetition and associate sequences from occurring in the same list.

The five resulting lists were recorded on cassettes and presented on a Sony TC-330 tape recorder. Ten subjects were assigned to each of the five lists. The interitem interval was 5 sec, with a warning tone 5 sec before the first word.

A practice task was used to familiarize subjects with the experimental task. The practice items were 35 three-digit numbers selected from a table of random numbers. A 50-item list was constructed, consisting of 20 once-presented numbers and 15 twice-presented numbers, the repeated numbers occurring at various lags. These were recorded and presented in the same manner as the experimental list, except that every subject heard the identical practice list.

Procedure

Before hearing the practice and experimental lists, the subjects heard a set of instructions emphasizing that on their response sheets they were to respond 'old' (plus) only to items that were repeated and 'new' (minus) to all other items. To prevent an encoding strategy emphasizing semantic associates, we did not inform the subjects of the presence of the associate pairs in the list. However, the subjects were told that no words would be presented more than twice and that about 150 of the words would be presented once, while the remaining 75 would be presented twice, for a total of 300 words. They were also told they should not go back to change any responses because they might change the wrong response or miss making the current response. Questions were answered after both the instructions and the practice list, and then the experimental list was presented without interruption.

RESULTS

As in previous studies using one session and single words as items, overall recognition performance was very high. Because the focus of this study was on the early portion of the false-alarm curve, relatively short lags were employed. Therefore, the hit curve had nearly no opportunity for any substantial decline (Raser, 1972). The proportion of hits for twice-presented items exceeded .9 at all systematic lags (for lags 0, 5, 10, and 15, the proportions were 1.00, .97, .94, and .97 respectively). Also, the mean proportion of false alarms to once-presented items was .09.

Turning to the results of major interest, the mean proportion of false alarms to the associates, collapsed over blocks and counterbalancing lists, is shown as a function of lag in Figure 2. The dashed line is the mean proportion of false alarms to the associate controls (only the second member of the pair presented) and shows the general rate of false alarms to these particular items. A repeated-measures analysis of variance on the false alarms to the associates demonstrated that the overall effect of lag was significant [$F(4, 196) = 2.78, MS_e = .01, p < .05$]. However, our primary concern is with the *form* of the lag function. In particular, if the nonmonotonicity of the lag function is reliable, then the quadratic trend over the four lags for the associates should be significant. Indeed, trend analysis did yield a significant quadratic component [$F(1, 196) = 6.00$, $p < .05$]; neither the linear component [$F(1, 196) = 1.50, p > .10$] nor the cubic component [$F(1, 196) = 1.00, p > .10$] was significant.

Another way of examining the lag function is to conduct a set of orthogonal comparisons of the associate controls versus the lag 0 associates, of the controls and lag 0 associates versus the lag 5, 10, and 15 associates, and

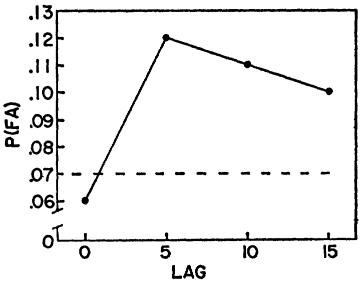

Figure 2. Proportion of false alarms to associates as a function of the lag between the prior item and its semantic associate. Also shown (dashed line) is the proportion of false alarms to the same items as associate controls (only the second member of the pair presented)

among the lag 5, 10, and 15 associates.[1] The results of this analysis showed no reliable difference between the controls and lag 0 associates [$p > .10$], a reliably higher rate of false alarms to the lag 5, 10, and 15 associates than to the controls and lag 0 associates [$p < .01$], and nonsignificant differences among the three longer lags [$p > .10$] although the latter showed the expected ordinal decrease in false alarms.

DISCUSSION

The magnitude of the effect of the presentation of the first member on false alarms to the second member of the associate pairs (i.e., false alarms to associates minus false alarms to associate controls) at lags 5, 10, and 15 was approximately 4% to 6%. This is in general agreement with the 6% to 10% found in previous studies that did not systematically vary the lag between semantic associates (e.g., Kimble, 1968; Zimmerman and Kimble, 1973).

The main finding, shown in Figure 2, is that the rate of false alarms to the associates did *not* monotonically decrease with lag. In fact, at lag 0, that rate did not even exceed the general rate of false alarms to the same items when they were semantically unrelated (i.e., the associate controls). These results disconfirm all of the previously mentioned one-step explanations of semantic generalization in recognition memory. They also disconfirm any overall explanation leading to a decision process based on a Thurstone scale such that at lag 0, the distribution for the new related items is higher in familiarity than that for the new unrelated items. What could account for these data is a process like that shown at the left of Figure 3, where the distribution for the new related items starts at that for the new unrelated items, subsequently shifts to the right, and finally shifts back to the left. We will briefly describe two different processes, each of which could yield such a result.

First, the circles at the center of Figure 3 illustrate a diffusion and decay of the trace element (Shepard, 1961). According to this view, generalization is a diffusion that does not occur immediately but rather during the lag (the dark arrow). Also, trace elements can decay out of both an old item and a new related item, as demonstrated by the dashed arrows. Thus, this is a *two-step* process (diffusion/decay) operating during the lag.

Second, the ovals at the far right of Figure 3 illustrate another process that would produce the nonmonotonic lag function. Each word to be remembered is assumed to be composed of two kinds of semantic features: general, G, and idiosyncratic, I, features (see Anisfeld and Knapp, 1968; Rips, Shoben, and Smith, 1973). In this 'G/I' model, the G features might, for example, correspond to information about taxonomic category,

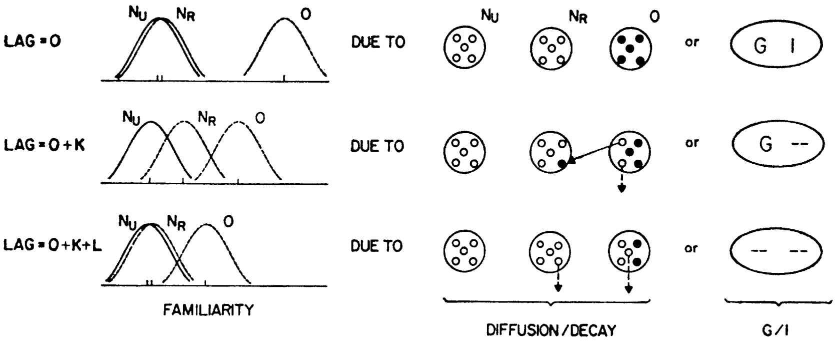

Figure 3. Comparison of three theoretical distributions of familiarity of the new unrelated (N_U), the new related (N_R), and the old (O) items at various study/test lags (0, 0 + K, 0 + K + L), as based on a diffusion/decay process or a G/I model of differential forgetting of general (G) and idiosyncratic (I) semantic features during the lag

while the I features might specify the particular exemplar from the category. During study, the G and I features of the target item (as well as the G features of items related to the target item) are activated. When the memory trace for the target item contains both G and I features, false alarms for activated items in the same category will be unlikely. However, when I features but not G features are lost, then false alarms for activated items in the same category will occur frequently. For example, if the memory trace for TABLE contains only G features (e.g., 'type of furniture'), then false alarms to CHAIR might occur. This notion accounts for the initial increase in the rate of false alarms for the new related words at short lags. And as the G features are lost, the rate of false alarms for the new related words should decrease toward that for the new unrelated words. We point out that the lags may need to be very long before the two rates approach each other: when Johnson (1970) used lags from 30 to 70, a lag of approximately 60 was necessary before the rate of false alarms to antonyms decreased to the rate of false alarms to semantically unrelated items. Finally, this model, also a *two-step* process, with separate forgetting of G and I features, may also help to account for what has been called class recognition as opposed to item recognition (see Kintsch, 1968).

One may question the need for such theorizing before the reliability of the nonmonotonic lag function has been established in other laboratories. Consider, for example, a recent study by Walter and Hellebusch (1974). One of their conclusions was that the lag function for semantically related false alarms is flat (i.e., neither nonmonotonic nor monotonic-decreasing).

However, their conclusion is based on difference scores (experimental minus control at each lag). Since their control items were words arbitrarily matched on several dimensions to the second member of the associate pair (rather than being the second members themselves, as in our study), the use of difference scores may be misleading. We reexamined their data on false alarms to *experimental* items only and found mean values as a function of lag of .18 (lag 0), .14 (lag 5), .28 (lag 10), .27 (lag 20), and .23 (lag 40), a marginally significant lag effect [$F(4, 76) = 2.40, MS_e = .03, .10 > p > .05$].[2] Acceptance of the null hypothesis (no lag effect) under such circumstances seems unwarranted, particularly since a nonmonotonic function has also been found for some kinds of nonsemantic similarity. Shepard and Chang (1963) and Batchelder and Johnson (1969) found results like ours for formally similar three-digit numbers; Raser (1972) found analogous results for acoustically similar words and orthographically similar words. Hence, these results seem widespread enough to call into serious question the currently popular one-step accounts of generalization effects in recognition memory. Notions about semantic generalization, whether expressed as implicit associative responses or as feature overlap, are in need of revision if they are to encompass the nonmonotonic lag function.

Notes

This research was supported by United States Public Health Service Grant MH-21037 to the second author. The paper was written while the authors were visitors at the University of California at Irvine in 1974/75. Reprint requests should be sent to Colin M. MacLeod, Department of Psychology, University of Washington, Seattle, Washington 98195. Received for publication May 5, 1975.

1. This set of orthogonal comparisons is, of course, not independent of the preceding trend analysis. We include it, however, because some readers may prefer to partition the data in this manner rather than in the manner defined by trend analysis.

2. Walter and Hellebusch (1974) examined both forward and backward associates; for purposes of comparison to our study, this reanalysis is of their forward associates only (e.g., TABLE-CHAIR). We thank Donald A. Walter for kindly providing us his data.

References

Anderson, J. R., and Bower, G. H. 1972. Recognition and retrieval in free recall. *Psychological Review* 79:97–123.

Anisfeld, M., and Knapp, M. 1968. Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology* 77:171–179.

Batchelder, W. H., and Johnson, R. A. 1969. A theoretical and empirical analysis of item similarity in continuous recognition memory. Paper read at the Michigan Mathematical Psychology meetings in August.

Bower, G. H. 1972. Stimulus-sampling theory of encoding variability. In *Coding processes in human memory,* ed. A. W. Melton and E. Martin. New York: Wiley.

Egan, J. P. 1958. Recognition memory and the operating characteristic. Indiana University Hearing and Communication Laboratory, AFCRC-TN-58-51, AD-152650.

Fillenbaum, S. 1969. Words as feature complexes: False recognition of antonyms and synonyms. *Journal of Experimental Psychology* 82:400–402.

Grossman, L., and Eagle, M. 1970. Synonymity, antonymity, and association in false recognition responses. *Journal of Experimental Psychology* 83:244–248.

Johnson, R. A. 1970. Recognition memory for antonym pairs as a function of the number of intervening trials. Master's thesis, University of Illinois at Urbana-Champaign.

Kimble, G. A. 1968. Mediating associations. *Journal of Experimental Psychology* 76:263–266.

Kintsch, W. 1968. Recognition and free recall of organized lists. *Journal of Experimental Psychology* 78:481–487.

Luce, R. D. 1963. Detection and recognition. In *Handbook of mathematical psychology,* vol. 1, ed. R. D. Luce, R. R. Bush, and E. Galanter. New York: Wiley.

Paivio, A., Yuille, J. C., and Madigan, S. A. 1968. Concreteness, imagery, and meaningfulness values of 925 nouns. *Journal of Experimental Psychology Monograph* 76(1, part 2).

Palermo, D. S., and Jenkins, J. J. 1964. *Word association norms: Grade school through college.* Minneapolis: University of Minnesota Press.

Raser, G. A. 1972. False recognition as a function of encoding dimension and lag. *Journal of Experimental Psychology* 93:333–337.

Rips, L. J., Shoben, E. J., and Smith, E. E. 1973. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior* 12:1–20.

Shepard, R. N. 1961. Application of a trace model to the retention of information in a recognition task. *Psychometrika* 26:185–203.

Shepard, R. N., and Chang, J. J. 1963. Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology* 65:94–102.

Shepard, R. N., and Teghtsoonian, M. 1961. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology* 62:302–309.

Underwood, B. J. 1965. False recognition by implicit verbal responses. *Journal of Experimental Psychology* 70:122–129.

Vogt, J., and Kimble, G. A. 1973. False recognition as a function of associative proximity. *Journal of Experimental Psychology* 99:143–145.

Walter, D. A., and Hellebusch, S. 1974. The role of associative elaboration in word recognition: Evidence for a two-stage test elaboration sequence. *Journal of Experimental Psychology* 103:79–84.

Zimmerman, J., and Kimble, G. A. 1973. Effects of incentive on false recognition. *Journal of Experimental Psychology* 97:264–266.