

False recognition as a function of lag and distinctiveness

G. WILLIAM HILL IV and S. DAVID LEONARD
University of Georgia, Athens, Georgia 30602

The false recognition effect in continuous recognition memory was examined as a function of lag, orthographic distinctiveness of early critical stimulus (CS) words and (presumably semantic) distinctiveness (D) of experimental (E) words. False recognition was not affected by orthographic distinctiveness, although both D level and lag had some effects. The results provided little support for a features encoding hypothesis (cf. MacLeod & Nelson, 1976) or for a strict implicit associative response interpretation. Possible explanations in terms of E word D level were discussed.

False recognition refers to differential false alarm rates for new items that are in some way related to earlier items on a list and false alarm rates for new items that are not so related. In a study that set the paradigm, Underwood (1965) demonstrated that false recognition of words in a continuous recognition memory task similar to that employed by Shepard and Teghtsoonian (1961) was a function of the associative relatedness between critical stimulus (CS) words shown early in the list and items shown later in the list, experimental (E) words. Underwood argued along with Bousfield, Whitmarsh, and Danick (1958) that two kinds of implicit responses are made by a subject to each verbal item, a representational response (RR), which is automatically made to an item as the process of perceiving it, and an implicit associative response (IAR), which is a function of the stimulus item. Thus, Underwood suggested that false recognition of E words was a function of a confusion between the RR produced by the E word and the IAR produced by the CS word.

Subsequent research, however, has not supported an imperative role of associative relatedness. Using synonyms as the critical pairs in a continuous recognition memory task, Anisfeld and Knapp (1968) demonstrated that a false recognition effect can be obtained solely on the basis of semantic similarity. They suggested that their results were inconsistent with the IAR theory. Anisfeld and Knapp's theory is based on the premise that each verbal item may be conceptualized as a complex of features including semantic (encompassing associative responses), syntactic, phonological, and orthographic characteristics. Each item is assumed to be encoded through selective activation of a number of

its salient features. The particular set of features that are encoded may vary. Thus, false recognition may occur because the "S may be led mistakenly to 'disregard' the distinguishing features and consider the two as identical" (Anisfeld & Knapp, 1968, p. 178).

Further, MacLeod and Nelson (1976) found magnitude of false recognition to be a nonmonotonic function of interitem lag, showing essentially no effect at Lag 0, maximal effect at Lag 5, and then decreasing effects. MacLeod and Nelson's explanation for this function was that at very short interitem intervals, both general and idiosyncratic semantic features of the word are remembered and the subject can distinguish between related items. The general features, however, tend to be retained longer than the idiosyncratic features. Thus, incorrect matches may occur on the basis of shared general features. The general features are eventually lost as the interval increases; thus, false alarm rate drops. Johnson (cited in MacLeod & Nelson, 1976), using longer interitem lags, found a similar relationship between lag and the level of false recognition, with essentially no false recognition at the extreme lag. Several other studies have suggested that the presence of an idiosyncratic feature may be a variable in recognition memory (Kausler, 1973; Kausler & Settle, 1975; Zechmeister, 1972). For example, Kausler and Settle (1975) found that homophone pairs of high orthographic distinctiveness (d) produced a reversal of the false recognition effect by facilitating correct rejection of the E words, while homophone pairs of low orthographic d still produced a false recognition effect.

Assuming MacLeod and Nelson's (1976) argument is correct, the magnitude of false recognition across interitem intervals could be manipulated by varying the strength of some idiosyncratic feature of the critical items. Some support for this notion has been provided by the results of Kausler and Settle (1975).

The present experiment was designed to compare the idiosyncratic features and IAR notions by using two

The report is based upon a master's thesis submitted to the Department of Psychology at the University of Georgia by the first author under the direction of the second author. Portions of the material presented here were presented to the 1977 meeting of the American Psychological Association. Requests for reprints should be sent to the second author.

separate procedures to manipulate the strength of the idiosyncratic features of the critical items. While the basic results of MacLeod and Nelson (1976) could as readily be explained by differential forgetting of the RR and IAR, given similar semantic characteristics but different orthographic characteristics, a shift in the point of maximal false recognition would be predicted by the feature-encoding hypothesis. Attempts were made to vary orthographic *d* without varying any semantic characteristics, and to vary (by selection) a more semantic characteristic of distinctiveness. While orthographic *d* within the list should not affect the IAR, it might be assumed to direct attention to a particular item resulting in more elaborate encoding, including its orthographic characteristics, thus enhancing both idiosyncratic and general features. If so, the interitem interval during which few false recognition responses occur should be extended due to better retention of the idiosyncratic features. An apparently semantic feature of the E words, distinctiveness was manipulated to examine the effect of strength of an idiosyncratic feature of E words on the magnitude of false recognition. The general and idiosyncratic features theory and the findings of Kausler and Settle (1975) suggest that with strong idiosyncratic features of the E word, the magnitude of false recognition should be lower regardless of the interitem interval. The IAR approach considers only strength of association to be relevant. Thus, idiosyncratic features of CS and E words would not affect false recognition.

The present experiment also included a larger range of interitem intervals than that employed by MacLeod and Nelson (1976) in an attempt to replicate the function obtained by Johnson (cited in MacLeod & Nelson, 1976), which found a return to the baseline level at extreme lags. It was assumed that the nonmonotonic lag function would be replicated with items having weaker idiosyncratic features (e.g., low in distinctiveness), but that the onset of false recognitions would be deferred with increasing distinctiveness.

METHOD

Subjects

A total of 123 undergraduate students at the University of Georgia participated in the experiment. The subjects were run in groups of two to eight with experimental list conditions randomly assigned except where necessary to complete a group.

During the course of the experiment, a number of subjects were noted to be making very few or no errors on critical items. In order to facilitate the subsequent analyses, a criterion of at least five errors on CS and/or E words was established. Thirty-eight subjects failed to meet the criterion, and five additional subjects were randomly discarded to equalize list conditions at eight.

Materials

From a previous continuous recognition memory experiment using two-syllable adjectives, a scale of D was defined as the number of hits minus number of false alarms. Several subjects generated synonyms to words drawn from the quartiles of this scale and selected to be relatively well matched on frequency

of usage and meaningfulness. The strength of synonymy of synonym pairs was then rated on a 5-point scale by a separate group of 47 subjects. The 10 highest rated pairs from each quartile of D were selected for the present experiment. The E words were all two-syllable adjectives scaled for D, while the CS words were all adjectives, but had not been scaled for D, frequency of usage or meaningfulness and did vary in number of syllables. The 20 control (C) words were two-syllable adjectives matched to the E words on the characteristics of D, meaningfulness, and frequency of usage. Two lists of 230 adjectives were composed from the 40 CS, 40 E, 20 C, and 30 filler words. Each filler word was repeated, as were 70 of the CS, E, and C words following the point at which the E or C item was tested.

All words were typed in capital letters and slides were produced. For each CS word, an additional slide was made in lowercase with a different typeface. The CS-E pairs were randomly divided into two sets of 20 pairs under the restriction of five E words from each D level appearing in each set. Five lists were derived, allowing each E word to occur at lags of 1, 5, 10, 20, and 40 items. Half the subjects receiving each list saw one of the CS sets in the orthographically distinctive type, while the other half saw the other set.

A postexperimental single-choice recognition test for the CS words, which included all 40 CS words and 40 distractors, was constructed in order to test the effectiveness of the isolation technique.

Procedure

The method of list presentation was similar to that employed by Underwood (1965). The items were presented visually at a 7-sec rate in a continuous recognition memory task. Subjects were instructed to decide whether an item was "old," meaning that it had appeared earlier on the list, or "new," meaning that the item had not previously appeared on the list. Subjects were given booklets in which to record their responses. Every 10th item was identified by the experimenter to help the subjects to maintain the correct position on the response sheets.

The subjects were instructed to attend to the words that were orthographically isolated on the list. They were warned, however, not to base their decisions on type size. They were not told of the separate recognition test for these items.

RESULTS

The overall performance of subjects was high, as had been noted in previous studies employing a continuous recognition memory task (e.g., MacLeod & Nelson, 1976), with the proportion of hits and correct rejections equal to .92 and .89, respectively. Analysis of the post hoc recognition test for isolated vs. nonisolated CS words showed no significant difference [$t(79) = .45$, $p < .25$]. However, since this task may have occurred after dissipation of the idiosyncratic features, the isolation variable was included in the other analyses.

The treatments by subjects design was deemed inappropriate for the analyses due to the limited range of errors. Therefore, a random-replications design was employed in the analyses of the false alarms to E and C words, with the five list orders treated as simple replications. A 2 by 3 by 4 by 5 analysis of variance (ANOVA) was performed with the particular half of the CS words that was orthographically isolated as a between-subjects (presumed dummy) factor and three levels of critical items (representing E words preceded by

isolated CS words, E words preceded by nonisolated CS words, and C words), four levels of D, and five levels of interitem lag as within-subjects factors.

Contrary to the prediction of the features hypothesis, the magnitude of false recognition was not significantly affected by CS presentation mode or even CS presentation [F(2,16) = 2.38, $p > .05$, MSe = 1.44]. The error means on E words by subject were 3.94, and 3.96 for orthographically distinct CS and standard CS presentation, respectively, and 3.54 for C words. The results also failed to replicate MacLeod and Nelson's (1976) finding that magnitude of false recognition was a function of interitem lags. Neither the main effect of lag [F(4,32) < 1] nor the Lag by Critical Item interaction [F(8,64) = 1.56, $p > .05$, MSe = 1.26] was significant.

Since the scale values of D were obtained from a highly similar task, it is not surprising that D level significantly affected the overall magnitude of false alarms for the critical items [F(3,24) = 38.30, $p < .001$, MSe = 1.48]. As expected, mean errors decreased with increasing levels of D (means = 3.89, 3.53, 2.74, and 1.29, from lowest to highest level). However, the D by Critical Item interaction was nonsignificant [F(6,84) = 2.05, $p > .05$, MSe = 1.08], as well as the D by Critical Item by Lag interaction [F(24,192) < 1].

While these results would best be termed negative with respect to the concept of false recognition generally, an unexpected significant interaction of Orthographic Set by Critical Item [F(2,16) = .29, $p < .01$, MSe = 1.44] was cause for a more detailed examination of the data. Inspection revealed that one of the sets of synonym pairs had produced a larger number of errors than either the other set of synonyms or the C words, although they were not differentially affected by presentation mode.

Based on this result, an analysis was performed on the errors to E and C words without respect to presentation mode but with E items divided into two sets. A 3 by 4 by 5 ANOVA was performed on the three sets of critical items (two E and one C), the four levels of D, and the five levels of lag as within-subjects factors. The critical item set effect in this analysis was significant [F(2,18) = 13.71, $p < .01$, MSe = 1.31], as was the effect of level D [F(3,27) = 40.78, $p < .001$, MSe = 1.39], although lag was again nonsignificant [F(4,36) < 1]. Examination of the simple effects of critical item group using the Newman-Keuls procedure showed one synonym group (the high-error group) produced significantly more false alarms than either the C words or the other synonym group (the low-error group). Thus, one could say the high-error synonym group produced a false recognition effect.

Further, this analysis produced a significant Lag by Critical Item interaction [F(8,72) = 2.09, $p < .05$, MSe = 1.21]. Post hoc comparisons of critical item groups at each lag indicated that while the low-error and C items were not significantly different at any of the lags,

the pattern of false recognition responses for the high-error E items in comparison with C items was essentially the same as obtained by MacLeod and Nelson (1976). The mean differences per replication between high-error E and C words were -.08, 3.3, 1.9, 2.4, and 1.8 for Lags 1, 5, 10, 20, and 40, respectively. That is, the minimal difference was found at Lag 1 and the maximal difference at Lag 5, with generally decreasing differences thereafter. With the exception of Lag 1, all differences between high-error E and C words were significant ($p < .05$). However, differences among lags for the high-error group only were not significant [F(4,32) = 1.79, $p > .10$, MSe = 2.22]. An analysis of the differences among lags based upon MacLeod and Nelson's prediction of a nonmonotonic function indicated that the difference between Lag 1 and Lag 5 was significant [$t(9) = 3.55$, $p < .01$], while the difference between Lag 5 and Lag 40 was not significant [$t(9) < 1$, $p > .10$].

The D by Critical Item interaction was also significant [F(6,54) = 8.59, $p < .01$, MSe = 1.07]. Although the mean number of false alarms decreased with increasing levels of D for all critical item sets, the high-error E words decreased much less than the low-error E words and C words. The differences among groups were not significant at the two lower levels of D but were at the two higher levels [F(2,18) = 33.89, $p < .01$, MSe = 3.42, and F(2,18) = 32.15, $p < .01$, MSe = 2.34, for D₃ and D₄, respectively]. Examination of simple effects showed high-error E words produced significantly more errors than either low-error words or C words. The D by Critical Item by Lag interaction, however, was not significant [F(24,192) < 1].

Since the E words in the high-error and low-error synonym groups were matched on D level, meaningfulness, and frequency of usage, it seemed unreasonable to presume that those items themselves differed enough conceptually to create the obtained difference. However, the CS words had not been equated with respect to these variables. Walter and Hellebusch (1974) had found evidence that CS words that produced false alarms themselves tended to produce fewer E word false alarms. They proposed that IARs were developed in searching for a memorial representation of the CS word when it was correctly classed as new. Contrarily, if the CS word was classed (incorrectly) as old, search would stop, fewer IARs would develop, and a false alarm to the E word would be less likely. Thus, the CS words associated with each of the synonym groups were examined for false alarms as well as the effect of orthographic d in an ANOVA that treated both as within-subjects factors. CS words for the high-error group had significantly fewer false alarms than CS words for the low-error group [F(1,78) = 32.07, $p < .001$, MSe = 2.76], with means of 2.15 and 3.64, respectively. But neither orthographic d [F(1,78) = 1.65, $p > .10$, MSe = 2.76] nor its interaction with the synonym sets [F(1,78) < 1] was significant.

Because of the differential false alarm rates to CS words for the high-error and low-error groups, it seemed analyses of the conditional probabilities of errors on the E words given errors on CS words were in order. Surprisingly, however, when individual words were examined, where "+" denotes false alarm, the $P(E+|CS+)$ was greater than the $P(E+|CS-)$ for 27 of the 40 E words. This would seem to be a statistically significant effect [$\chi^2(1) = 4.90, p < .05$].

DISCUSSION

The results of the present experiment are somewhat complicated by the failure of one set of synonyms to show a false recognition effect. At present there are no obvious differences among either CS or E words to account for this. However, based on the extant literature, we must assume that the high-error synonym set is representative of what occurs when synonyms are presented in a task of this sort. Unfortunately, the data provide no strong support for either a strict IAR position or a feature-encoding approach.

The overall pattern of differences between the high-error E word set and the C words was similar to that obtained by MacLeod and Nelson (1976). However, failure to find a significant trend of return to baseline at long lags is at variance with the results they reported for Johnson's (cited in MacLeod & Nelson, 1976) work. While this study used synonyms, recall of the general features of the CS word should not be dependent on whether or not an antonym or a synonym follows it. Although it could be argued that synonyms share more general features, one should expect some difference between lags of 5 and 40. Moreover, while D level reduced the false recognition effect, the failure to obtain a D by Critical Item by Lag interaction or an effect of orthographic d suggests that if idiosyncratic features protect against false recognition, their effect is very short-lived.

The generally greater probability of false alarms to CS words in the low-error synonym set seemed to support Walter and Hellebusch's (1974) hypothesis that more IARs are generated when the CS is correctly recognized as new. However, explicit analysis by word produced an opposite finding.

The fact that D level produced the most dependable results (essentially a replication) suggests the possibility that the two-stage approach that may be most useful is one that emphasizes the characteristics of the search process as well as the storage process. Since subjects are asked to make a determination of new as well as old, we must assume that they make some tests that examine some features of the present as well as the stored items. These features might include but would not be limited to IARs. Different words vary on the extent to which rejection may readily be made. This is reflected in the D level of the

words. At low D levels a relatively high rate of false alarms occur, and in the present study no differences were obtained among the three sets of words. As D levels increased, false alarms to the low-error and C word sets decreased dramatically, while the decrease in the high-error set was much smaller. Since D level was varied by matching, it is conceivable that poor matching occasioned this result. However, MacLeod and Nelson (1976) got a false recognition effect using the same items as E and C words. Although they did not control for D, one might assume their E words varied in D and the differential effect on high-D words could account for the overall significance.

It seems that synonymity alone is not adequate to interfere with the rejection characteristics of high-D words; for example, the low-error set of synonyms was unaffected. In addition, physical changes in the stimuli seemed to have no effect. It is possible that associative factors other than synonymity may have contributed to the effect. One also cannot rule out possible effects of phonological, syntactic, and other features. It would appear that future research should vary these characteristics concomitantly while controlling for D level.

REFERENCES

- ANISFELD, M., & KNAPP, M. Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology*, 1968, **77**, 171-179.
- BOUSFIELD, W. A., WHITMARSH, G. A., & DANICK, J. J. Partial response identities in verbal generalization. *Psychological Reports*, 1958, **4**, 703-713.
- KAUSLER, D. H. Orthographic distinctiveness of homonyms and the feature-tagging hypothesis. *American Journal of Psychology*, 1973, **86**, 141-149.
- KAUSLER, D. H., & SETTLE, A. V. The false-recognition effect for homophones of items subjects are cued to forget. *American Journal of Psychology*, 1975, **88**, 627-634.
- MACLEOD, C. M., & NELSON, T. O. A nonmonotonic lag function for false alarms to associates. *American Journal of Psychology*, 1976, **89**, 127-135.
- SHEPARD, R. N., & TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302-309.
- UNDERWOOD, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, **70**, 122-129.
- WALTER, D. A., & HELLEBUSCH, S. The role of associative elaboration in word recognition: Evidence for a two-stage test elaboration sequence. *Journal of Experimental Psychology*, 1974, **103**, 79-84.
- ZECHMEISTER, E. B. Orthographic distinctiveness as a variable in word recognition. *American Journal of Psychology*, 1972, **85**, 425-430.

(Received for publication January 31, 1979.)