

Errors in a recognition memory task are a U-shaped function of word frequency

EUGENE B. ZECHMEISTER, CAROL CURT, and JULIE A. SEBASTIAN
Loyola University, Chicago, Illinois 60626

College students rated verbal items for either familiarity or meaningfulness and were then tested for their recognition memory of rated items from four critical classes: nonwords and words appearing less than 1 time, 1-10, and more than 40 times per million words of written text (Thorndike-Lorge, 1944). Following either rating task, recognition errors were found to be a U-shaped function of word frequency and nonwords were recognized least well. Although a U-shaped function was expected on the basis of what is known about word frequency effects in general, there is significant disagreement as to the mechanism(s) responsible for this relationship.

Memory for low-frequency (LF) words is generally superior to that for high-frequency (HF) words when recognition is tested. This word frequency effect (WFE) is quite reliable, having been shown, for example, for both written and aural presentation, with paced and unpaced study, and in absolute-judgment as well as in forced-choice testing procedures (Schulman, 1967; Shepard, 1967; Underwood & Freund, 1970). However, notably missing among studies investigating the WFE is any systematic manipulation of word frequency across a wide range of frequencies, the WFE typically being reported when recognition for words of only two rather disparate levels of background frequency has been tested. Yet, there is speculation that probability of correct recognition is an "inverted-U-shaped" (Schulman, 1976, p. 307) or "curvilinear" (Gregg, 1976, p. 213) function of frequency, were recognition to be examined over the complete range of word frequencies.

Schulman (1976) reached this conclusion after observing college students' recognition memory for very rare words they previously had rated for familiarity. He found that probability of correct recognition increased with an item's rated familiarity, showing a "reversal" of the usual WFE. As noted, generally only two classes of words are used to show the typical WFE and, since these items almost always can be described as "very familiar" and "relatively unfamiliar" (but not totally unfamiliar), on the basis of his results Schulman reasonably predicted that probability of correct recognition should first increase and then decrease as familiarity (frequency) increased.

Gregg (1976) also suggests a nonmonotonic relationship between recognition probability and background frequency. He points to the findings of Allen and Garton (1968), who "manipulated" word familiarity

by having physics and arts students view both physics and common words. Overall, physics words were recognized better than common words, demonstrating the WFE. However, physics students recognized physics words better than did arts students, while there was little difference between the two groups of students when common words were tested. If we think of the arts student studying physics words as somehow similar to the average psychology subject studying very rare words, these results also indicate that recognition memory must get better before it gets worse as background frequency (familiarity) increases. As will be shown, for the materials chosen and the conditions used in the present experiment, these investigators are correct in their expectations.

METHOD

Lists

A study list consisted of 120 items, with 36 of these items later tested for recognition (critical items) and the remaining items serving as fillers. In constructing a list, 33 words were first randomly selected from each of three Thorndike and Lorge (1944) frequency levels: Items appearing more than 40 times, 1 through 10 times, and less than 1 time per million. In addition, 33 nonwords were chosen from several sources or constructed by the authors, with the restriction that the items corresponded to English orthography. Nine items from each of the four classes were randomly designated as critical items, 15 as fillers, and 9 items to be used as new items on the recognition test. Finally, 24 more filler words were randomly selected from those words in Thorndike and Lorge listed as occurring 11 through 40 times per million (12) or less than 1 time per million (12). Therefore, half the items in the study list had frequencies greater than 1 and half were words with frequencies of less than 1 or were nonwords; among words with frequencies greater than 1, there were no "gaps" in the frequency range. Items were assigned randomly to positions in the study list such that an equal number of items from each verbal class appeared in the first and second halves of the list.

The recognition test involved a forced-choice procedure, with critical items paired with new items of the same verbal class. Either four or five critical items from each of the four verbal classes were tested in the first half of the test list; the others

Requests for reprints should be addressed to Eugene B. Zechmeister, Department of Psychology, Loyola University of Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626.

were tested in the second half. Furthermore, approximately half of the first-half test items had appeared in the first half of the study list and half had appeared in the second half.

In the experiment two study lists were used equally often in each experimental condition. In the two lists, filler items were the same, but the 36 critical items in one list served as new items on the recognition test for the other list, and vice versa.

Procedure

Both the study list and recognition test were presented in booklet form. The cover page announced the research as a "Word Scaling Study" and a following page explained that the purpose of the study was to examine persons' knowledge of various words. No mention was made that a memory test would be given. One half of the subjects received instructions asking them to rate verbal items on a 3-point meaningfulness scale: 1, definitely know the meaning as it might be found in a dictionary; 2, not sure, maybe know the dictionary meaning; 3, definitely do not know the meaning. The other half of the subjects received booklets asking them to rate the items for familiarity: 1, very familiar, seen it many times before; 2, not very familiar, seen it but not very frequently; 3, not at all familiar, never seen this item. The 3-point scales were reprinted on half sheets of paper and the subjects were asked to keep the scales in front of them while doing the rating. Instructions included the information that the items to be rated were both words (including those easily recognized and those not so easily recognized) and nonwords. The 120 items appeared in four columns of 30 items on four consecutive booklet pages. The task was self-paced, with ratings made in the space next to each item. Subjects were told not to look ahead or back in their booklets once they started to rate the items.

The last page of the booklet was labeled "Recognition Test," and 36 pairs of items appeared in two columns of 18 pairs. Instructions explained that one item in each pair was from the rating list and this item was to be circled on the sheet.

Subjects

Booklets were distributed in a block randomization procedure, in which each block contained one each of four different booklets (2 forms of the study lists by 2 sets of instructions). A total of 88 subjects participated. Testing was done in small groups and all subjects were students from introductory psychology classes, with participation earning extra credit in the course.

RESULTS

Proportion recognition errors for each of the four classes of verbal items and for each instruction condition is shown in Figure 1. Errors were reliably different among verbal classes [$F(3,252) = 8.46, p < .01$] and, although the students consistently made fewer errors for all classes of items when meaning as opposed to familiarity was rated, this difference was not significant. Also, the interaction between instructions and item classes was not significant, nor were there any interactions with list forms.

A Newman-Keuls test carried out on the mean errors made in each of the four classes of verbal items (combining instruction groups) showed no difference between mean errors for very rare words ($< 1/\text{million}$) and common words ($> 40/\text{million}$); however, all other comparisons were significant ($p < .05$). Clearly, recognition errors are a U-shaped function of word frequency,

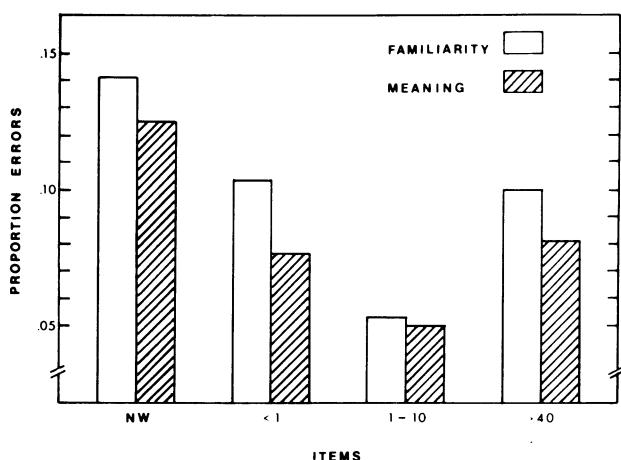


Figure 1. Proportion recognition errors made by two groups of college students after rating verbal items for either familiarity or meaning. The items were nonwords (NW) and words appearing less than 1 time, 1 through 10 times, and more than 40 times per million in the Thorndike-Lorge count.

and, when all classes of items are considered, errors are greatest for nonwords (NW).

As must be expected, in nearly all instances common words received a rating of 1, very familiar or very meaningful (99%), and nonwords received a rating of 3, unfamiliar or not meaningful (94%). However, very rare words were rated sufficient numbers of times at each scale point to allow recognition to be examined as a function of the rating on the combined meaningfulness and familiarity scales. For these items proportion correct recognition for each of the three scale points, 3, 2, and 1, was: .89 (428/481), .91 (130/143), and .97 (163/168). That probability of correct recognition increases as rated familiarity or meaningfulness of rare words increases agrees with the previously cited report by Schulman (1976). Although items appearing 1-10 times per million were also differentially rated on both scales, a similar relationship between probability of recognition and rating did not emerge. In terms of increasing familiarity or meaningfulness, the proportions were: .95 (94/99), .94 (135/143), and .95 (519/546). It should be noted that recognition was extremely good for this class of items.

DISCUSSION

The frequency theory of recognition memory (cf. Underwood & Freund, 1970) accounts for the WFE on the basis of implicit associative responses (IARs). According to the theory, IARs are more likely to occur to HF than LF study words, and these implicit responses serve to augment the frequency of HF new words, producing a more difficult discrimination for this class of items. Obviously, very rare words, which in the present experiment "behaved" much like HF words, would not be susceptible to the same type of interference and, hence, the frequency theory must posit an additional mechanism to account for the obtained U-shaped function. One possibility is that very rare words or nonsense words produce fragmented representa-

tional responses. If frequency accrues to individual parts of an item, this may be a source of false recognition when these fragments are found in new items, as has been shown for multisyllable test words constructed from syllables that had not appeared together in study words (Underwood & Zimmerman, 1973).

It is also possible that the "processing" of uncommon (but not totally unfamiliar) words is different from that of very rare or of very familiar words. For example, Ghatala and her associates (Ghatala & Levin, 1974; Ghatala, Levin, & Makoid, 1975) have shown that LF words have an advantage over HF words in both frequency judgment tasks and verbal discrimination tasks, but only when the subjects (in this case, children) know the meaning of the LF words. Ghatala et al., for example, compared (Experiment 1) children's discrimination learning of HF words with LF words that were either of high meaningfulness (more than 80% of the children were able to give a definition) or of low meaningfulness (less than 20% could give a definition). Verbal discrimination learning was better for LF words which the children could also define than for LF words which they could not define or for HF words. According to Ghatala and Levin (1974), items that are both low in background frequency and have meaning for the subjects "produce larger frequency 'units' and/or more stable 'units'" (p. 399). Ghatala, Levin, and Wilder (1975) have further suggested that greater stability of frequency units is related to the depth of processing of items.

Although not expressed in frequency terms, Lockhart, Craik, and Jacoby (1976) have also suggested that greater depth of processing accompanies LF words and that this explains the usual WFE. According to these authors, while the processing of common words involves a minimum of processing operations, that of uncommon words requires more elaborate and deeper operations, thereby providing a more unique or distinctive encoding. Both the frequency-processing argument of Ghatala and the distinctiveness-processing explanation of Lockhart et al. can be said to be supported by the present study, given the assumption that both very rare words and very common words are not processed as deeply as uncommon words. In fact, evidence for such an assumption can be found in a recent experiment by Kolars (1976). When college students were asked whether they knew certain words well enough to use them in a sentence, reaction times revealed "a sort of U-shaped function between speed of responding and frequency of occurrence of the item in the language" (p. 555). Whether "depth" or "complexity of operations" is a proper accounting of these results is unclear to us; however, Kolars' finding does suggest that uncommon words require more cognitive effort of some kind.

If uncommon (but known) LF words require greater and/or more complex processing, there appears to be yet another potential advantage in recognition memory for these words which results, not from processing per se, but from the fact that the cognitive activities associated with processing an item are likely to be encoded as part of the memory for that word.

We have made a similar argument to explain the superior recognition found for structurally unusual words (Zechmeister, 1972) or for words which subjects have been instructed to "process" in a specific manner (Zechmeister & Gude, 1974).

REFERENCES

- ALLEN, L. R., & GARTON, R. F. The influence of word-knowledge on the word-frequency effect in recognition memory. *Psychonomic Science*, 1968, 10, 401-402.
- GHATALA, E. S., & LEVIN, J. R. Discrimination learning as a function of differences in materials: A proposed explanation. *Memory & Cognition*, 1974, 2, 395-400.
- GHATALA, E. S., LEVIN, J. R., & MAKOID, L. A. A clarification of frequency effects in children's discrimination learning. *Memory & Cognition*, 1975, 3, 1-6.
- GHATALA, E. S., LEVIN, J. R., & WILDER, L. Pronunciation and the frequency-meaningfulness effect in children's frequency discrimination. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 1, 655-659.
- GREGG, V. Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. New York: Wiley, 1976.
- KOLERS, P. A. Knowing not. *Memory & Cognition*, 1976, 4, 553-558.
- LOCKHART, R. S., CRAIK, F. I. M., & JACOBY, L. Depth of processing, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. New York: Wiley, 1976.
- SCHULMAN, A. I. Word length and rarity in recognition memory. *Psychonomic Science*, 1967, 9, 211-212.
- SCHULMAN, A. I. Memory for rare words previously rated for familiarity. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 301-307.
- SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 156-163.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University Teachers College, 1944.
- UNDERWOOD, B. J., & FREUND, J. S. Word frequency and short-term recognition memory. *American Journal of Psychology*, 1970, 83, 343-351.
- UNDERWOOD, B. J., & ZIMMERMAN, J. The syllable as a source of error in multisyllable word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 701-706.
- ZECHMEISTER, E. B. Orthographic distinctiveness as a variable in word recognition. *American Journal of Psychology*, 1972, 85, 425-430.
- ZECHMEISTER, E. B., & GUDE, C. Instruction effects in recognition memory. *Bulletin of the Psychonomic Society*, 1974, 3, 13-15.

(Received for publication February 16, 1978.)