JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **13**, 63–69 (1974)

The Mechanism of the Word-Frequency Effect on Recognition Memory[1]

MARCEL KINSBOURNE AND JAMES GEORGE

Duke University

The better recognition memory for rare as compared to frequently used words was shown to be unrelated to difference in association value, imagery, concreteness, or any change in phenomenal frequency due to exposure on the test list. But a manipulation substantially modified recognition scores. Words that were exposed just prior to the learning task for ostensibly unrelated purposes were much less well recognized than ones not so exposed. The frequency-recognition effect was thus simulated in the laboratory by manipulating situational frequency.

Frequently used words are better recalled than words that are rarely used (Bousfield & Cohen, 1955; Hall, 1955; Sumby, 1963). This effect is hypothesized to be largely due to the higher mean association value of the frequent words (Deese, 1960). In contrast, it is the less frequent words that are easier to recognize in an ensemble (Allen & Garton, 1968; Clayton & Evarts, 1972; Gorman, 1961; McCormack & Swenson, 1972; Schulman, 1967; Schwartz & Rouse, 1961; Shepard, 1967; Underwood, 1972; Underwood & Freund, 1970). Some have attempted to explain the effect by relying on supposedly correlated variables such as association value. This is higher for frequent words and is assumed to interfere in recognition by contaminating memory with task-irrelevant associates (Allen & Garton, 1968; Schulman, 1967; Underwood & Freund, 1970). Any such role of concreteness was ruled out by Gorman (1961). Some have invoked proactive inhibition, from which rare words are said to be relatively immune (Gorman, 1961); McCormack & Swenson, 1972; Shepard, 1967). Other have suggested that rare words are in some manner structurally distinct and thus more easily recognized (McCormack & Swenson, 1972; Schulman, 1967; Zechmeister, 1972).

All these hypotheses assume equal effort on the part of subjects in initially memorizing rare and frequent words. One cannot be sure of this on the basis of the reports in the literature, but it is at least plausible when the lists for study included words from only one frequency band. When lists mixed as regards word frequency are used, as by Shepard (1967) and Clayton and Evarts (1972), then rare words may attract more attention and memorizing effort than common words. The possibility of differing amount of rehearsal of rare and frequent words within the same list makes the results of these experiments hard to interpret.

Rare words are indeed somewhat different structurally from frequent words (Kinsbourne & Evens, 1970; Landauer & Streeter, 1973), but neither this nor proactive inhibition will be considered in the first experiment. Instead, we just evaluate the notion that the frequency effect is mediated by correlated differences in association value. We also assess one possible mechanism for the effect of frequency. The latter is as follows. A rare word, exposed in a study list, by virtue of this additional exposure becomes effectively more frequent for the subject. When subsequently presented in an

[1] We acknowledge the help of John Marslen-Wilson in running the first experiment. This work was supported in part by Public Health Service Research Grant HD00668 from the N.I.C.H.D.

Copyright © 1974 by Academic Press, Inc.
All rights of reproduction in any form reserved.
Printed in Great Britain

ensemble, it stands out from the other rare words by virtue of its ostensibly higher frequency, and this disparity guides the subject's selection. For frequent words, a similar effect would obtain, but it would be less marked, as an already frequent word increases only slightly in ostensible frequency on reexposure (Noble, 1959).

To evaluate the role of association value, it was varied independently of word frequency. As regards the "higher frequency" bestowed by test exposure, the recognition ensembles were constructed so as to deprive rare words of the benefit of that hypothesized change. Each ensemble included a representative of each of nine wavebands along the frequency spectrum (the high and low frequencies representing the extremes of the spectrum). If test exposure moves a rare word into a higher frequency bracket, this would make the word no less confusable, as there would still be in the ensemble some word of comparable frequency from which it would need to be discriminated.

A control recall condition was included in the experimental design.

Experiment I

Method

Subjects: The subjects were 96 student volunteers from a College of Further Education in Oxford, England. Their ages ranged from 16 to 23, with a mean of 18.6 years. All subjects were naive with respect to any form of psychological experiment.

Materials. Four lists of 16 words each were constructed, with interitem (intralist) association, word frequency, and method of testing being the main variables. Thus, there was one High Frequency High Association (HFHA) list, one High Frequency Low Association (HFLA) list, one Low Frequency High Association (LFHA) list, and one Low Frequency Low association (LFLA) list. Forty-eight subjects were tested on all four lists by a free recall method and 48 by a recognition method.

As published association data in the English language are based on American norms, the source used for interitem association was the association tables for English young adults compiled by Kinsbourne and Evans (1967). The HA lists each consisted of eight highly associated word-pairs taken from these tables, with association percentages ranging from 23 % to 82 %. The mean frequency of association was held constant for both lists. Since there were some further associations between items, apart from those between the original pairs, parity was maintained across the two lists for this as well. The two LA lists consisted of words selected at random, within the given frequency limits, which, according to the association tables, were unassociated. The words within each list were randomized.

Word frequency was derived from the L-count in Thorndike and Lorge's *Teacher's Wordbook of 30,000 Words* (1944). The frequency bands for the lists were arbitrarily selected as 1–65 for the LF lists and 1000+ for the HF lists (divide by 4.5 to derive the number of appearances per one million words). Thus, all the words used on the two LF lists scored between 1 and 65 on the L-count, with both lists having the same mean word frequency . The same was true for the two HF lists, where all words scored 1000 or above.

The lists were thus balanced for association value and for frequency and also for syllable length—each list containing 10 one-syllable and six two-syllable words.

In the free recall condition, the subjects were given forms with 16 blanks to fill in. In the recognition condition, the subjects were presented with recognition arrays. Four arrays were compiled, one for each list, and each consisting of 16 blocks of 10 words, in which were embedded the 16 stimulus words, one to each block. Further, in making up the arrays, the spectrum of word frequencies was divided up into 10 bands, two for the list LF band, two for the list HF band, with the remaining spectrum (66–999) divided into six equal bands, as follows:

Band Number	*L-count Frequency*
1	1–34
2	35–65
3	66–219
4	220–375
5	376–531
6	532–687
7	688–843
8	844–999
9	1000–2500
10	2500+

Each array of 10 words thus consisted of one word from each of the 10 bands, with care being taken that the stimulus word should replace a word of the same frequency as itself. The same balance was maintained between the one- and two-syllable words as in the stimulus lists.

Design and procedure. There were four groups of 24, two male and two female. One group of each sex was tested by recall and the other by recognition. A 2 × 2 design was used, rotating association value and frequency, so that within each group of 24 there were four subgroups of six learning the lists in different orders, so as to allow for possible practice or fatigue effects.

The subjects were told they would be asked to learn a number of lists and were shown a sample list. The lists were typed in capitals in six, five, and five words on slips of paper. They would be allowed 35 sec in which to learn the lists. The method of answering was then explained and samples of the two sheets shown. In the recall condition, they were to write down as many of the words on the lists as they could remember. In the recognition condition, they were to underline one word in each of the 16 blocks, corresponding to the 16 words on the lists. If they could not remember the word in any particular block, they were required to guess. The subjects were allowed unlimited time for answering (although in practice, 5 min was usually more than enough time). The whole procedure took no more than 30–40 min.

Results

Mean scores (out of 16 possible correct) and standard deviations are given in Table 1.

TABLE 1

MEAN NUMBER OF CORRECTLY RECALLED WORDS (OUT OF 16) UNDER THE CONDITIONS OF EXPERIMENT I

	Frequency	
Associations	High	Low
Recall		
High	11.8	11.2
	(2.3)	(2.8)
Low	10.1	9.3
	(2.4)	(2.2)
Recognition		
High	13.2	14.8
	(2.4)	(1.5)
Low	(11.5)	(13.6)
	(3.1)	(2.3)

Analysis of variance revealed significant main effects for response mode, $F(1,70) = 102.50$, $p < .001$, association value, $F(1,70) = 37.26$, $p < .001$, and frequency $F(1,70) = 4.90$, $p < .05$. The only significant interaction was response mode × frequency, $F(1,70) = 24.44$ $p < .001$. Thus, unlike association value, frequency differently affects the relative efficiency of recall and recognition.

The percentage distribution of recognition errors over frequency bands 1 (lowest) through 10 (highest) is presented in Table 2, separately for the high-frequency lists (HF) and the low-frequency lists (LF). Inspection showed that association value had no effect on this distribution so the lists of different frequencies were collapsed across association value. A combined percentage is given for bands 9 and 10 for HF, and 1 and 2 for LF, because in the respective lists one or the other of these bands contained the correct target. Thus, opportunities for error on these bands were half as great as on the rest. It was predicted that false recognitions would tend to cluster on frequency bands comparable to that of the test list. Spearman rank order correlations between the predicted and actual distribution of errors across frequency bands were significant both for HF (rho = .80, $p < .02$) and LF (rho = .87, $p < .01$) lists.

TABLE 2

THE PERCENTAGE DISTRIBUTION OF RECOGNITION ERRORS OVER FREQUENCY BANDS

	Band number									
Lists	1	2	3	4	5	6	7	8	9	10
HF	3.8	4.6	6.8	10.2	18.3	13.1	11.7	16.3	15.2	
LF	18.2		13.7	13.9	11.9	8.2	8.3	12.9	7.7	5.2

Discussion

The findings in the recall condition confirm Deese (1960), in that high association value favored recall. It had a similar effect on recognition, suggesting that the ability to organize a word list can help recognition as well as recall, in spite of previous failures to find such an effect (Cofer, 1967; Kintsch, 1968). The difference may depend on the nature of the distractor items, and in particular, on their associative relationship with the test words.

The opposing effect of word frequency on recall and on recognition was confirmed. It here occurred with use of mixed frequency ensembles. Previously it had been demonstrated with ensembles of words matched for frequency with the words on the test list. The more frequent words were better recalled but less well recognized. This effect was distinct from and not in the same direction as the effect of high association value, which enhanced both recall and recognition. It could not be attributed to differential distribution of attention between rare and frequent words while memorizing, as the frequency band was held constant within a list. The advantage in recognition for rare words could not be due to the previously viewed rare word having assumed an effectively higher frequency due to the additional exposure and thus standing out from the other low-frequency words in the recognition list, because the recognition lists covered the spectrum of frequency between that designated "high" and that designated "low" in this study. A shift in effective frequency would have left the target word as confusable as before, irrespective of the degree of shift.

When subjects made mistakes, they did not randomly choose distractor words irrespective of their frequency. Rather, after studying high-frequency word lists, subjects' errors tended to involve the choice of high-frequency distractors. After studying low-frequency lists, they tended to make their false recognitions among low-frequency distractor words. This suggests that subjects gained some implicit sense of the prevalent word frequency within the list they were studying, and this affected their guesses at times when they were uncertain in their selection of the previously displayed word from the recognition ensemble.

We conclude that the effect of frequency on word recognition must be due to some aspect of the frequency of exposure itself rather than to differential study of the list words or differences due to different association value. The frequency effect is not due to enhanced salience of the target word within an equal frequency recognition list due to a change in its effective frequency after experimental exposure, as a range of frequencies were represented in each ensemble. Direct confirmation of the effectiveness of the frequency variable awaits the manipulation of that variable itself. If such a manipulation substantially affects ease of recognition, then possibly correlated variables, such as supposedly structural distinctiveness or rare words, need not be used as an explanatory principle. The second experiment is an attempt at such a manipulation.

Experiment II

A direct manipulation of the subject's experience with the words to be remembered is attempted in order to explore further the locus of the frequency effect.

Method

Subjects. Forty students from Duke University introductory psychology courses participated for credit.

Materials. The following types of word lists were constructed:

1. Four memory lists consisting of 16 words apiece.

2. Four recognition sets made up of 16 lists of 10 words. A recognition set contains all 16 words from the memory list with one target word per list.

3. Two preexposure lists each consisting of all the words from one of the two recognition sets (160) plus 160 fill-in words.

The words for the memory lists were randomly selected from Paivio, Yuille, and Madigan's (1968) norms for 925 nouns, with the following restrictions. Two lists were made up of high-frequency words (over 50 per million on the Thorndike-Lorge count, and the other two consisted of low-frequency words (under 20 per million). The lists were controlled for imagery, concreteness, and meaningfulness by choosing lists with an equal average value on each of these factors.

The words for the recognition sets were randomly chosen from Thorndike and Lorge (1944). The only restriction on these lists was the requirement that the sets be in the same frequency range as the target words (i.e., 1–20 per million for two sets and over 50 per million for the other two).

The 160 fill-in words for each preexposure list were again chosen from Thorndike and Lorge's lists, with the only requirement being that they were not words chosen in (1) or (2) above. Each exposure list also contained one high-frequency and one low-frequency recognition set.

The length of words was balanced across the memory lists and recognition sets (most words had two syllables, and the rest three).

Design and procedure. Subjects were run two at a time. They were first asked to rate the concreteness of the words on one preexposure list. Half were given one preexposure list, half the other list, in order to counter-balance the effect of particular word lists. This task increased the subjects' recent experience with the words that were later given on two of the four memory lists. The subjects were told that this task was not related to the upcoming memory task.

In the second part of the experiment, the subjects were given 32 sec per list to memorize the words on each of the four memory lists. They were told to try to pay equal attention to each of the words. Each memory list was followed by its recognition set after an unfilled 30-sec delay. The subjects were told to pick the word from each of the 16 lists that was the same as one on the memory lists and to guess if they were unsure. The lists were presented in four different orders (six subjects per order).

Results

The effects of three factors on recognition memory were analyzed. The first was the frequency of the words which was either low or high. The second was the immediate previous experience with the words which were either exposed on the rating task or not exposed. The third was the order that the subject received the lists; there were four levels of this factor. Thus, the results were analyzed using a 4 (order) × 2 (frequency) × 2 (exposure) design with repeated measures on frequency and exposure. Mean scores (out of 16 possible correct) and standard deviations are shown in Table 3. They are collapsed across the order factor since it did not prove to be significant. As can be seen, even with the attempt to hold all other factors constant, there is a strong frequency effect, $F(1,36) = 36.03$, $p < .001$. There is also a strong effect of immediate past experience with the words, $F(1,36) = 20.09$, $p < .001$. The interaction was not significant, $F < 1$. A preplanned specific contrast between the low-frequency exposed and the high-frequency nonexposed conditions was nonsignificant, but a contrast between the low-frequency exposed and the high-frequency exposed conditions was significant, $F(1,36) = 22.3$, $p < .001$.

TABLE 3

MEAN SCORES (OUT OF 16 POSSIBLE CORRECT) AND STANDARD DEVIATIONS

	Frequency	
Preexposure	High	Low
Exposed	8.7 (3.0)	11.2 (3.1)
Not exposed	10.3 (3.1)	12.8 (2.5)

Discussion

The results of the second experiment again demonstrate the frequency effect on word-recognition memory. The effect occurred although imagery and concreteness were held constant between the rare and frequent word sets. Further, those words which had been exposed to the subject immediately before the learning task for ostensibly unrelated reasons and in a manner that did not call for any effort to memorize were ultimately less well recognized than those that were not preexposed. The magnitude of this effect was such that preexposed rare words were no better recognized than frequent words that had not been preexposed. Preexposed frequent words were least well recognized of all.

This result renders all accounts based on supposedly correlated variables or on supposedly great distinctiveness of rare words unnecessary. A single preexposure was sufficient to cause a major decline in recognition scores.

Thus, a change in situational frequency affected recognition memory in a manner that simulated the frequency-recognition effect. By what mechanism did this occur?

Without preexposure, subjects discriminate between zero previous intraexperimental occurrences (of distractors) and a single previous occurrence (of target items). With preexposure, the discrimination is more difficult, namely, between one appearance (of distractors) as against two (of targets). But a recency account can also be given. In both conditions, the study list words are more recent than the distractors, but preexposure drastically reduces the difference between these two degrees of recency. In all previous demonstrations of the frequency-recognition effect frequency and recency are similarly confounded, though extraexperimentally rather than situationally. However, a further aspect of the results renders a simple recency model untenable. The preexposure manipulation equalizes situational recency of extra-experimentally frequent and rare words. Nevertheless, preexposed frequent words were significantly less well recognized than preexposed rare words, there being no significant interaction between the frequency effect and the effect of preexposure. Thus, cumulative frequency of previous experience had an effect over and above that which could be accounted for by difference in time elapsed since the most recent viewing of a given word.

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.

BOUSFIELD, W. A., & COHEN, B. M. The occurrence of clustering in the recall of randomly arranged words of different frequencies of usage. *Journal of Genetic Psychology*, 1955, **52**, 83–95.

CLAYTON, K. N., & EVARTS, M. Recognition memory and the word frequency effect. Paper presented to the Southeastern Psychological Association, Atlanta, 1972.

COFER, C. N. Does conceptual organization influence the amount retained in immediate free recall? In B. T. Kleinmuntz (Ed.), *Concepts and the structure of memory*, New York: Wiley, 1967.

DEESE, J. Frequency of usage and number of words in free recall: the role of association. *Psychological Reports*, 1960, **7**, 337–366.

GORMAN, A. N. Recognition memory for names as a function of abstractness and frequency. *Journal of Experimental Psychology*, 1961, **61**, 23–29.

HALL, J. F. Learning as a function of word frequency. *American Journal of Psychology*, 1955, **67**, 138–160.

KINSBOURNE, M., & EVANS, R. Comparative word association norms for children, young adult and elderly samples. Unpublished manuscript, 1967.

KINSBOURNE, M., & EVENS, N. Is the word frequency effect on recognition threshold a function of transitional probability? *Psychological Reports*, 1970, **27**, 143–145.

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

LANDAUER, T. K., & STREETER, L. A. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 119–131.

MCCORMACK, P. D., & SWENSON, A. L. Recognition memory for common and rare words. *Journal of Experimental Psychology*, 1972, **95**, 72–77.

NOBLE, C. E. The familiarity-frequency relationship. *Journal of Experimental Psychology*, 1959, **67**, 13–16.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76**, 1–25.

SCHULMAN, A. I. Word length and rarity in recognition memory. *Psychonomic Science*, 1967, **9**, 211–212.

SCHWARTZ, F., & ROUSE, R. D. The activation and recovery of associations. *Psychological Issues*, 1961, **3**, 1–40.

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

SUMBY, W. M. Word frequency and serial position effect. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 443–450.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 Words*. New York: New York Bureau of Publication, Teachers College, Columbia University, 1944.

UNDERWOOD, B. J. Word recognition memory and frequency information. *Journal of Experimental Psychology*, 1972, **94**, 276–283.

UNDERWOOD, B. J., & FREUND, J. J. Testing effects in the recognition of words. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 117–125.

ZECHMEISTER, E. B. Orthographic distinctiveness as a variable in word recognition. *American Journal of Psychology*, 1972, **85**, 425–430.

(Received June 6, 1973)