



Published in final edited form as:

J Exp Psychol Learn Mem Cogn. 2013 November ; 39(6): 1943–1946. doi:10.1037/a0033669.

Parametric effects of word frequency effect in memory for mixed frequency lists

Lynn J. Lohnas and Michael J. Kahana

University of Pennsylvania

Abstract

The word frequency paradox refers to the finding that low frequency words are better recognized than high frequency words yet high frequency words are better recalled than low frequency words. Rather than comparing separate groups of low and high frequency words, we sought to quantify the functional relation between word frequency and memory performance across the broad range of frequencies typically used in episodic memory experiments. Here we report that both low frequency and high frequency words are better recalled than mid-frequency words. In contrast, we only observe a low frequency advantage when participants were given a subsequent item recognition test. The U-shaped relation between word frequency and recall probability may help to explain inconsistent results in studies using mixed lists with separate groups of high and low frequency words.

Introduction

In item recognition tasks, low frequency (i.e., rare) words are more easily recognized as targets and more easily rejected as lures than high frequency (i.e., common) words (Gorman, 1961). In free recall tasks, lists of high frequency words are generally better recalled than lists of low frequency words (Hall, 1954; Sumbly, 1963). These twin findings have been termed the word frequency paradox, and a variety of theories have been suggested to account for these findings (Coane, Balota, Dolan, & Jacoby, 2011; Criss & Malmberg, 2008; Dennis & Humphreys, 2001; Gillund & Shiffrin, 1984; Glanzer, Adams, Iverson, & Kim, 1993; Gregg, 1976; Heathcote, Ditton, & Mitchell, 2006; Maddox & Estes, 1997; Malmberg & Murnane, 2002; Malmberg & Nelson, 2003; McDaniel & Bugg, 2008; Reder et al., 2000; Shepard, 1967; Shiffrin & Steyvers, 1997; Steyvers & Malmberg, 2003). The low frequency advantage in item recognition is upheld in lists comprised of both low and high frequency items (Criss & Malmberg, 2008; Dorfman & Glanzer, 1988; Estes & Maddox, 2002; Glanzer & Adams, 1985; Gorman, 1961; Heathcote et al., 2006; Malmberg, Steyvers, Stephens, & Shiffrin, 2002; Shepard, 1967). However, in such mixed lists the superior recall of high frequency words is less robust: Some mixed list experiments exhibit better recall of low frequency words (DeLosh & McDaniel, 1996; Merritt, DeLosh, & McDaniel, 2006; Ozubko & Joordens, 2007), some exhibit better recall of high frequency words (Balota & Neely, 1980; Hicks, Marsh, & Cook, 2005), and some exhibit no reliable difference between

low and high frequency words (May, Cuddy, & Norton, 1979; Ozubko & Joordens, 2007; Ward, Woodward, Stevens, & Stinson, 2003; Watkins, LeCompte, & Kim, 2000).

We found the present state of affairs unsettling. Given the robust advantage of low-frequency items in recognition memory, why is the effect seemingly unstable in free recall of mixed lists? We suggest that this instability arises from the experimental manipulation commonly used to examine these effects: an “extreme groups design” compares memory for two groups of words that differ substantially in their range of word frequencies. Yet, if in mixed lists, recall favors low frequency words (as in item recognition) in addition to favoring high frequency words, such an experimental design cannot inform whether there are simultaneous recall advantages for both low and high frequency words. We address these issues by characterizing the functional relation between word frequency and recall performance in mixed frequency lists in which word frequencies varies continuously across a broad range. We also examine the word frequency effect on a final recognition test of the words presented in the recall task. We consider whether these effects are modulated by the presence of an encoding task.

Methods

The data reported in this manuscript were collected as part the Penn Electrophysiology of Encoding and Retrieval Study, involving three multi-session experiments that were sequentially administered. Here we include 132 participants (age 17–30, mean = 22.1 ± .3) who have completed the first phase of the experiment. These participants consisted of students and staff at the University of Pennsylvania, Drexel University, Rowan University, Temple University, University of the Arts, and the University of the Sciences.

Each of seven sessions consisted of 16 lists of 16 words presented one at a time on a computer screen. Each study list was followed by an immediate free recall test and each session ended with a recognition test. Half of the sessions were randomly chosen to include a final free recall test which took place before the recognition test.

Each word was drawn from a pool of 1638 words (available at http://memory.psych.upenn.edu/files/wordpools/PEERS_wordpool.zip). Each item was on the screen for 3000 ms, followed by jittered 800 – 1200 ms inter-stimulus interval. Words were either presented concurrently with a task cue, indicating that a participant should make one of two encoding judgments for that word and indicate their response via keypress, or with no encoding task. The two encoding tasks were a size judgment (“Will this item fit into a shoebox?”) and an animacy judgment (“Does this word refer to something living or not living?”), and the current task was indicated by the color and typeface of the presented item. Using the results of a prior norming study, only words that were clear in meaning and that could be reliably judged in the size and animacy encoding tasks were included in the pool. There were three types of lists: no-task lists (subjects did not have to perform judgments with the presented items), single-task lists (all items were presented with the same task), and task-shift lists (both types of judgments were used in a list, although each item was presented with only one judgment type). Here we only distinguish task lists from no-task

lists, as our primary focus is the influence of a semantic encoding task on memory performance.

After the last item in the list, there was a 1200 – 1400 ms jittered delay, after which a tone sounded, a row of asterisks appeared, and the participant was given 75 s to attempt to recall any of the just-presented items. If a session was randomly selected for final free recall, following the immediate free recall test from the last list, participants were shown an instruction screen for final free recall, telling them to recall all the items from the preceding lists. After a 5 s delay, a tone sounded and a row of asterisks appeared. Participants had 5 minutes to recall any item from the preceding lists.

A recognition test was administered after either final free recall or the last list's immediate recall test. In this final recognition test, lures were selected from the remaining 1638 items not presented during the free recall phase, and target/lure ratio varied with session, where targets made up 80, 75, 62.5, or 50 percent of the total items. In total, 320 words were presented one at a time on the computer screen. When a word was presented on the screen, participants were instructed to indicate whether the test word had been presented previously. Participants were told to respond verbally “pess” for old items and “po” for new items and to confirm their response by pressing the space bar. These responses (“pess” and “po”) were chosen that both response types would initiate with the same stop consonant (or plosive) so as to assist in automated detection of word onset times. Following the old/new judgment, participants made a confidence rating on a scale of 1 to 5, with 5 being the most confident. Recognition was self-paced though participants were encouraged to respond as quickly as possible without sacrificing accuracy. Participants were given feedback on accuracy and reaction time.

Because we report a post-hoc analysis of previously collected data, our original choice of words was not specifically designed to address questions of word frequency. Of the 1638 words used in our study, we included in our analyses the 984 words for which we could obtain imageability and concreteness measures in the MRC database (Wilson, 1988). For each of these words, we obtained an estimate of the frequency of usage in the English language using the CELEX2 database (Baayen, Piepenbrock, & Gulikers, 1995) which defines frequency as counts per million in the Birmingham corpus (Sinclair, 1987). The word pool was then partitioned into 10 approximately equally-sized bins ranging from low to high frequency counts (see Table 1). Because some words shared the same frequency value, the bins could not be exactly the same size, but each bin contained between 9.3% and 10.6% of possible frequency values. Across word frequency bins, one-way ANOVAs for each of concreteness, imageability, and word length revealed that the words did not vary in any of these dimensions across frequency bins (all $p > 0.05$).

Results

Figure 1 shows a striking U-shaped relation between word frequency and recall irrespective of whether items were presented with an encoding task. In a 10×2 repeated-measures ANOVA with recall probability as the dependent variable, and frequency bin and the presence of an encoding task as factors, we found both main effects to be significant

(frequency bin: $F(9, 2489) = 15.2, p < 0.001$; task presence: $F(1, 2489) = 129, p < 0.001$). There was also a significant interaction between frequency and task ($F(9, 2489) = 3.08, p < 0.005$). To ensure that the effect of word frequency was significant in both task types, we performed repeated-measures ANOVAs separately for each of the encoding task types. For both of these ANOVAs, the main effect of frequency bin is still significant (no task: $F(9, 1179) = 3.75, p < 0.001$; task: $F(9, 1179) = 24.8, p < 0.001$).

To assess the recall advantage for low frequency and high frequency words, we defined low frequency words as those in the lowest bin and high frequency as the highest bin; mid-frequency words comprised the remaining eight frequency bins. Recall of low frequency words and high frequency words were significantly higher than recall of mid-frequency words (low versus medium, no task: $t(131) = 2.16, p < 0.05$; low versus medium, task: $t(131) = 3.52, p < 0.001$; high versus medium, no task: $t(131) = 3.03, p < 0.005$; high versus medium, task: $t(131) = 12.3, p < 0.001$).

Figure 2 shows that on a subsequent item recognition task, a monotonic effect of word frequency is observed for targets and lures (Criss & Malmberg, 2008; Estes & Maddox, 2002). In a two-factor repeated-measures ANOVA with hit rate as the dependent variable all effects were significant (frequency bin: $F(9, 2489) = 34.6, p < 0.001$; task presence: $F(1, 2489) = 15.4, p < 0.001$; interaction: $F(9, 2489) = 2.04, p < 0.05$). As with recall probability, one-way repeated measures separately based on the presence of an encoding task still yielded a significant effect of frequency bin (no task: $F(9, 1179) = 17.4, p < 0.001$; task: $F(9, 1179) = 22.9, p < 0.001$). In addition, a one-way repeated-measures ANOVA with false alarm rate as the dependent variable and frequency bin as the factor (as lures did not have associated encoding tasks) revealed a significant main effect ($F(9, 1179) = 47.7, p < 0.001$).

Participants exhibit lower false alarm rates for low frequency than mid-frequency words ($t(131) = 14.2, p < 0.00001$) and for mid-frequency than high frequency words ($t(131) = 4.58, p < 0.001$). The hit rates for targets are higher for low versus mid-frequency targets (no task: $t(131) = 7.66, p < 0.001$; task: $t(131) = 5.60, p < 0.001$) as well as for mid-frequency vs. high frequency targets (no task: $t(131) = 6.33, p < 0.001$; task: $t(131) = 7.42, p < 0.001$).

Discussion

By examining a wide range of frequencies in mixed lists we have found significant benefits of both low and high frequency words in recall. Our analysis of word-frequency effects demonstrates a clear U-shaped pattern in free recall, favoring recall of both low and high frequency words over mid-frequency words. We find the expected low frequency word advantage in item recognition for hit rates and false alarm rates. Each of these effects was present both for freely-encoded items and for items encoded while participants made a size or animacy judgment task.

The nonmonotonic word frequency effect shown in Figure 1 may help to explain the inconsistent results obtained in previous studies that limited comparisons to distinct categories of low and high frequency words. In our data set, a comparison of recall

performance for the lowest and highest word frequency bins would suggest an advantage for high frequency words. One could imagine that different definitions of low frequency and high frequency could lead to comparisons of different bins of items in Figure 1, which could lead to a low frequency advantage, high frequency advantage, or no difference in performance as a function of word frequency.

Although one might find it tempting to comment on the inconsistent findings in prior research of free recall and word frequency, we hesitate to reinterpret previous findings derived from studies that relied on comparisons between groups of low and high frequency words. Furthermore, our parametric U-shaped relation frequency and recall does not speak directly to previous work showing that intentionality of encoding (Watkins et al., 2000) and the temporal ordering of low and high frequency words (Ozubko & Joordens, 2007) may interact with the degree to which high and low frequency items are favored in recall. Nonetheless, the present findings of a non-monotonic word frequency effect illustrate the importance of considering frequency as a continuous rather than a dichotomous variable in evaluating theoretical accounts of how frequency interacts with performance in recall and recognition tasks.

Acknowledgments

The authors gratefully acknowledge support from National Institutes of Health grant MH55687. We thank Jonathan Miller and Patrick Crutchley for assistance with designing and programming the experiment, and we thank Kylie Hower, Joel Kuhn, and Elizabeth Crutchley for help with data collection.

References

- Baayen, R.H.; Piepenbrock, R.; Gulikers, L. The CELEX lexical database. [CD-ROM]. Philadelphia, PA: Linguistic data consortium. CD-ROM; 1995.
- Balota DA, Neely JH. Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning & Memory*. 1980; 6(5):576–587.
- Coane JH, Balota DA, Dolan PO, Jacoby LL. Not all sources of familiarity are created equal: the case of word frequency and repetition in episodic recognition. *Memory & Cognition*. 2011; 39:791–805. [PubMed: 21264634]
- Criss AH, Malmberg KJ. Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency. *Journal of Memory and Language*. 2008; 59:331–345.
- DeLosh EL, McDaniel MA. The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1996; 22(5):1136–1146.
- Dennis S, Humphreys MS. A context noise model of episodic word recognition. *Psychological Review*. 2001; 108:452–478. [PubMed: 11381837]
- Dorfman D, Glanzer M. List composition effects in lexical decision and recognition memory. *Journal of Memory and Language*. 1988; 27:633–648.
- Estes WK, Maddox WT. On the processes underlying stimulus-familiarity effects in recognition of words and nonwords. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 2002; 28(6):1003–1018.
- Gillund G, Shiffrin RM. A retrieval model for both recognition and recall. *Psychological Review*. 1984; 91:1–67. [PubMed: 6571421]
- Glanzer M, Adams JK. The mirror effect in recognition memory. *Memory & Cognition*. 1985; 13(1): 8–20. [PubMed: 4010518]

- Glanzer M, Adams JK, Iverson G, Kim K. The regularities of recognition memory. *Psychological Review*. 1993 Jul; 100(3):546–567. [PubMed: 8356189]
- Gorman AM. Recognition memory for nouns as a function of abstractedness and frequency. *Journal of Experimental Psychology*. 1961; 61:23–39. [PubMed: 13707272]
- Gregg, V. Word frequency, recognition and recall. In: Brown, J., editor. *Recall and recognition*. Oxford, England: John Wiley and Sons; 1976.
- Hall J. Learning as a function of word-frequency. *American Journal of Psychology*. 1954; 67:138–140. [PubMed: 13138778]
- Heathcote A, Ditton E, Mitchell K. Word frequency and word likeness mirror effects in episodic recognition memory. *Memory & Cognition*. 2006; 34(4):826–838. [PubMed: 17063914]
- Hicks JL, Marsh RL, Cook GI. An observation on the role of context variability in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31(5):1160–1164.
- Ku era, H.; Francis, W. *Computational analysis of present-day American English*. Providence, RI: Brown University Press; 1967.
- Maddox WT, Estes WK. Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1997; 23:539–559.
- Malmberg KJ, Murnane K. List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28(4):616–630.
- Malmberg KJ, Nelson TO. The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*. 2003; 31:35–43. [PubMed: 12699141]
- Malmberg KJ, Steyvers M, Stephens JD, Shiffrin RM. Feature frequency effects in recognition memory. *Memory & Cognition*. 2002; 30(4):607–613. [PubMed: 12184562]
- May RB, Cuddy LJ, Norton JM. Temporal contrast and the word frequency effect. *Canadian Journal of Psychology*. 1979; 33(3):141–147. [PubMed: 519544]
- McDaniel MA, Bugg JM. Instability in memory phenomena: a common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*. 2008; 15(2):237–255. [PubMed: 18488637]
- Merritt PS, DeLosh EL, McDaniel MA. Effects of word frequency on individual-item and serial order retention: Tests of the order-encoding view. *Memory & Cognition*. 2006; 34(8):1615–1627. [PubMed: 17489288]
- Ozubko J, Joordens S. The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review*. 2007; 14(5):871–876. [PubMed: 18087952]
- Reder LM, Nhouyvanisvong A, Schunn CD, Ayers MS, Angstadt R, Hiraki KA. A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2000; 26:294–320.
- Shepard RN. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*. 1967; 6:156–163.
- Shiffrin RM, Steyvers M. A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin and Review*. 1997; 4:145. [PubMed: 21331823]
- Sinclair, J., editor. *Looking up*. London/Glasgow: Collins/COBUILD; 1987.
- Steyvers M, Malmberg KJ. The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2003; 29(5):760–6.
- Sumby WH. Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior*. 1963; 1:443–450.
- Ward G, Woodward G, Stevens A, Stinson C. Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2003; 29:186–210.
- Watkins MJ, LeCompte DC, Kim K. Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2000; 26:239–245.
- Wilson MD. The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavior Research Methods, Instruments & Computers*. 1988; 20(1):6–11.

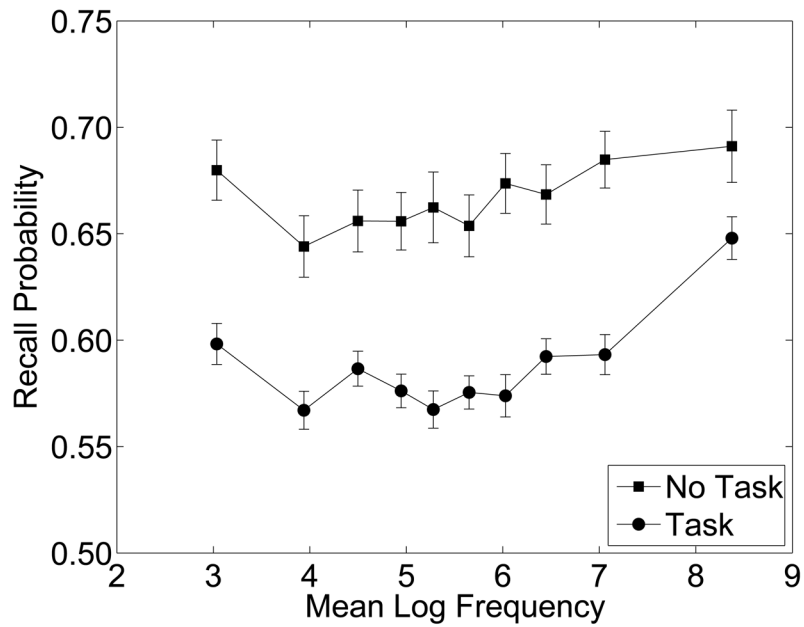


Figure 1. Word frequency effect in free recall

Participants recalled higher proportions of both low frequency and high frequency words than words of intermediate frequency, irrespective of whether the item was presented without an encoding task (filled squares) or with an encoding task (filled circles). The 984 included in this analysis were partitioned into deciles on the basis of their word frequency counts in the CELEX2 database. Each point corresponds to the mean recall probability for a decile of word frequencies.

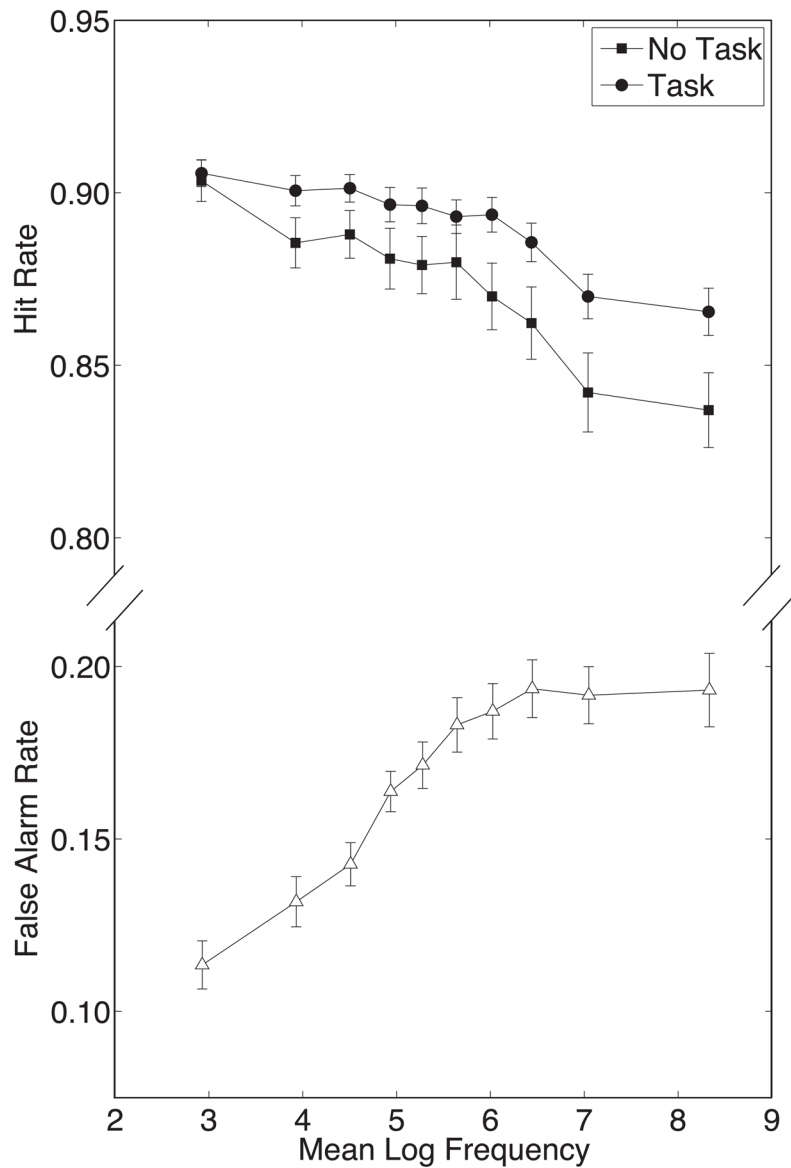


Figure 2. Word frequency effect in a post-recall item recognition test

Participants were more likely to incorrectly accept lures with increasing word frequency (open symbols), and less likely to correctly recognize targets with increasing word frequency (filled symbols), irrespective of whether the items were presented with an associated encoding task (circles) or no task (squares). The 984 included in this analysis were partitioned into deciles on the basis of their word frequency counts in the CELEX2 database. Each point corresponds to the mean recognition response for one word frequency decile.

Table 1

Frequency information for each word bin, quantified as counts per million in the Birmingham corpus (Sinclair, 1987), as provided in the CELEX2 database (Baayen et al., 1995).

Bin	Range	Mean
1	2–36	21
2	37–68	51
3	69–115	90
4	116–163	141
5	165–235	196
6	237–344	285
7	345–495	415
8	496–816	632
9	829–1575	1163
10	1589–26215	4332