Journal of Experimental Psychology:
Learning, Memory, and Cognition
2013, Vol. 39, No. 6, 1943–1946

0278-7393/13/$12.00 DOI: 10.1037/a0033669

RESEARCH REPORT

Parametric Effects of Word Frequency in Memory for Mixed Frequency Lists

Lynn J. Lohnas and Michael J. Kahana
University of Pennsylvania

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 midfrequency 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.

Keywords: recall, recognition, word frequency effect

In item recognition tasks, low frequency (i.e., rare) words are more easily recognized as targets and more easily rejected as lures than are 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; Sumby, 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 comprising 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 find 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 memory for two groups of words that differ substantially in their range of word frequencies. 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 indicate 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 vary 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.

Method

The data reported in this article were collected as part the Penn Electrophysiology of Encoding and Retrieval Study, involving three multisession experiments that were sequentially administered. Here we include 132 participants (ages 17–30 years, M = 22.1 years, SD = 0.3) who have completed the first phase of the experiment. These participants consisted of students and staff at the University of Pennsylvania, Drexel University, Rowan Uni-

This article was published Online First July 8, 2013.

Lynn J. Lohnas and Michael J. Kahana, Department of Psychology, University of Pennsylvania.

We 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.

Correspondence concerning this article should be addressed to Michael J. Kahana, 3401 Walnut Street, Suite 316C, Philadelphia, PA 19104. E-mail: kahana@psych.upenn.edu

versity, 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 that took place before the recognition test.

Each word was drawn from a pool of 1,638 words (available at http://memory.psych.upenn.edu/). Each item was on the screen for 3,000 ms, followed by a jittered 800- to 1,200-ms interstimulus 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 presented 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, we included only words that were clear in meaning and that could be reliably judged in the size and animacy encoding tasks in the pool. There were three types of lists: no-task lists (participants 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 1,200- to 1,400-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 min 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 1,638 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% 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 so that both response types would initiate with the same stop consonant (or plosive), thus assisting 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, although 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 1,638 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 analyses of variance (ANOVAs) for concreteness, imageability, and word length revealed that the words did not vary in any of these dimensions across frequency bins (all $p > .05$).

Results

Figure 1 shows a 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: For frequency bin, $F(9, 2489) = 15.2$, $p < .001$; for task presence, $F(1, 2489) = 129$, $p < .001$. There was also a significant interaction between frequency and task, $F(9, 2489) = 3.08$, $p < .005$. 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: For no task, $F(9, 1179) = 3.75$, $p < .001$; for task, $F(9, 1179) = 24.8$, $p < .001$.

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 words as those in the highest bin; midfrequency words comprised the remaining eight frequency bins. Recall of low frequency words and high frequency were significantly higher than recall of midfrequency words: For low versus medium, no task, $t(131) = 2.16$, $p < .05$; for low versus medium, task, $t(131) = 3.52$, $p < .001$; for high versus medium, no task, $t(131) = 3.03$, $p < .005$; for high versus medium, task, $t(131) = 12.3$, $p < .001$).

Table 1
Frequency Information for Each Word Bin

Bin	Range	*M*
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	829–1,575	1,163
10	1,589–26,215	4,332

Note. The frequency for each word bin is quantified as counts per million in the Birmingham Corpus (Sinclair, 1987), as provided in the CELEX2 database (Baayen, Piepenbrock, & Gulikers, 1995).

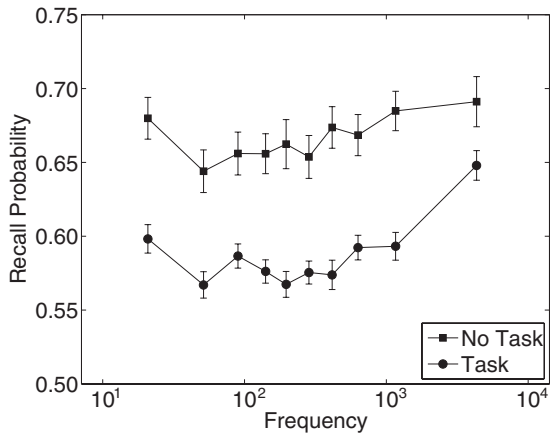

Figure 1. Word frequency effect in free recall. Participants recalled higher proportions of both low frequency and high frequency words than words of intermediate frequency, irrespective of whether the item was presented without an encoding task (filled squares) or with an encoding task (filled circles). 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. Error bars represent 95% Loftus and Masson (1994) confidence intervals.

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: For frequency bin, $F(9, 2489) = 34.6$, $p < .001$; for task presence, $F(1, 2489) = 15.4$, $p < .001$; for interaction, $F(9, 2489) = 2.04$, $p < .05$. 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: For no task, $F(9, 1179) = 17.4$, $p < .001$; for task, $F(9, 1179) = 22.9$, $p < .001$. 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, $F(9, 1179) = 47.7$, $p < .001$.

Participants exhibit lower false alarm rates for low frequency than midfrequency words, $t(131) = 14.2$, $p < .00001$, and for midfrequency than high frequency words, $t(131) = 4.58$, $p < .001$. The hit rates for targets are higher for low versus midfrequency targets: For no task, $t(131) = 7.66$, $p < .001$; for task, $t(131) = 5.60$, $p < .001$. This also held for midfrequency versus high frequency targets: For no task, $t(131) = 6.33$, $p < .001$; for task, $t(131) = 7.42$, $p < .001$.

Discussion

By examining a wide range of frequencies in mixed lists, we found significant benefits of both low and high frequency words in recall. Our analysis of word-frequency effects demonstrates a U-shaped pattern in free recall, favoring recall of both low and high frequency words over midfrequency 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.

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

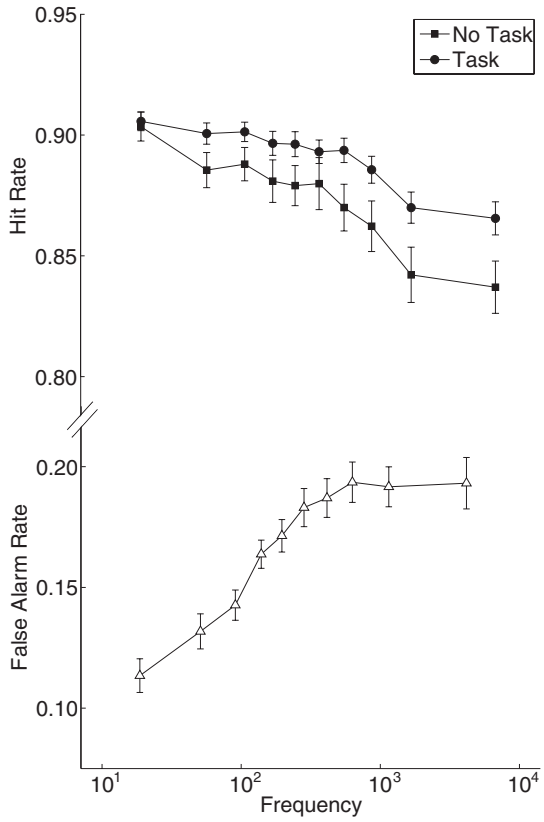

Figure 2. Word frequency effect in a postrecall item recognition test. Participants were more likely to incorrectly accept lures with increasing word frequency (open symbols) and less likely to correctly recognize targets with increasing word frequency (filled symbols), irrespective of whether the items 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. Error bars represent 95% Loftus and Masson (1994) confidence intervals.

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 nonmonotonic word frequency effect illustrate the importance of considering frequency a continuous rather than a dichotomous variable in evaluating theoretical accounts of how frequency interacts with performance in recall and recognition tasks.

References

Baayen, R. H., Piepenrbrock, R., & Gulikers, L. (1995). *The CELEX lexical database.* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 576–587. doi:10.1037/0278-7393.6.5.576

Coane, J. H., Balota, D. A., Dolan, P. O., & Jacoby, L. L. (2011). Not all sources of familiarity are created equal: The case of word frequency and repetition in episodic recognition. *Memory & Cognition, 39,* 791–805. doi:10.3758/s13421-010-0069-5

Criss, A. H., & Malmberg, K. J. (2008). Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency. *Journal of Memory and Language, 59,* 331–345. doi:10.1016/j.jml.2008.05.002

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1136–1146. doi:10.1037/0278-7393.22.5.1136

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478. doi:10.1037/0033-295X.108.2.452

Dorfman, D., & Glanzer, M. (1988). List composition effects in lexical decision and recognition memory. *Journal of Memory and Language, 27,* 633–648. doi:10.1016/0749-596X(88)90012-5

Estes, W. K., & Maddox, W. T. (2002). On the processes underlying stimulus-familiarity effects in recognition of words and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1003–1018. doi:10.1037/0278-7393.28.6.1003

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67. doi:10.1037/0033-295X.91.1.1

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20. doi:10.3758/BF03198438

Glanzer, M., Adams, J. K., Iverson, G., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567. doi:10.1037/0033-295X.100.3.546

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractedness and frequency. *Journal of Experimental Psychology, 61,* 23–39.

Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. Oxford, England: Wiley.

Hall, J. (1954). Learning as a function of word-frequency. *American Journal of Psychology, 67,* 138–140. doi:10.2307/1418080

Heathcote, A., Ditton, E., & Mitchell, K. (2006). Word frequency and word likeness mirror effects in episodic recognition memory. *Memory & Cognition, 34,* 826–838. doi:10.3758/BF03193430

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). An observation on the role of context variability in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1160–1164. doi:10.1037/0278-7393.31.5.1160

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490. doi:10.3758/BF03210951

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 539–559. doi:10.1037/0278-7393.23.3.539

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 616–630. doi:10.1037/0278-7393.28.4.616

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition, 31,* 35–43. doi:10.3758/BF03196080

Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition, 30,* 607–613. doi:10.3758/BF03194962

May, R. B., Cuddy, L. J., & Norton, J. M. (1979). Temporal contrast and the word frequency effect. *Canadian Journal of Psychology, 33,* 141–147. doi:10.1037/h0081712

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15,* 237–255. doi:10.3758/PBR.15.2.237

Merritt, P. S., DeLosh, E. L., & McDaniel, M. A. (2006). Effects of word frequency on individual-item and serial order retention: Tests of the order-encoding view. *Memory & Cognition, 34,* 1615–1627. doi:10.3758/BF03195924

Ozubko, J. D., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review, 14,* 871–876. doi:10.3758/BF03194114

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. A. (2000). 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, 26,* 294–320. doi:10.1037/0278-7393.26.2.294

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163. doi:10.1016/S0022-5371(67)80067-7

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166. doi:10.3758/BF03209391

Sinclair, J. (Ed.). (1987). *Looking up: An account of the COBUILD Project in lexical computing*. London and Glasgow, United Kingdom: Collins.

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 760–766. doi:10.1037/0278-7393.29.5.760

Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior, 1,* 443–450.

Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 186–210. doi:10.1037/0278-7393.29.2.186

Watkins, M. J., LeCompte, D. C., & Kim, K. (2000). Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 239–245. doi:10.1037/0278-7393.26.1.239

Wilson, M. (1988). The MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments & Computers, 20,* 6–10.

Received March 11, 2013
Revision received April 29, 2013
Accepted May 2, 2013 ■

See page 1725 for a correction to this article.

Correction to Lohnas and Kahana (2013)

In the article "Parametric Effects of Word Frequency in Memory for Mixed Frequency Lists" by Lynn J. Lohnas and Michael J. Kahana (*Journal of Experimental Psychology: Learning, Memory, and Cognition*, Advance online publication. July 8, 2013. doi:10.1037/a0033669) there were omissions in Figure 1. All versions of this article have been corrected.

DOI: 10.1037/a0034164