Journal of Experimental Psychology: Learning, Memory, and Cognition

The Shape of Things to Come: Evaluating Word Frequency as a Continuous Variable in Recognition Memory

Pernille Hemmer and Amy H. Criss

Online First Publication, July 15, 2013. doi: 10.1037/a0033744

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/13/$12.00 DOI: 10.1037/a0033744

RESEARCH REPORT

The Shape of Things to Come: Evaluating Word Frequency as a Continuous Variable in Recognition Memory

Pernille Hemmer
Rutgers University

Amy H. Criss
Syracuse University

The role of experience in memory, specifically the word frequency (WF) mirror effect showing higher hit rates and lower false alarm rates for low-frequency words, is one of the hallmarks of memory. However, this "regularity of memory" is limited because normative WF has been treated as discrete (low vs. high). We evaluate the extent to which the prototypical WF mirror effect holds when WF is treated as a continuous variable. We find a clear nonmonotonic U-shaped relationship. Hit rates are higher for *both* low-frequency and high-frequency words. Linear and quadratic regression models were fit to the data at both the item and the participant level, and the quadratic model provided a better fit at both levels. This finding is inconsistent with the empirical and theoretical finding of a mirror effect and requires a novel approach to accounting for the role of experience in episodic memory.

Keywords: episodic memory, word recognition, mirror effect, memory models, recognition memory

Word frequency (WF) has received much empirical and theoretical attention (e.g., Glanzer & Adams, 1985, 1990; Glanzer & Bowles, 1976; Schulman, 1967). In the laboratory WF is defined as the frequency of occurrence that a word appears in a corpus. For example, the word *pencil* has high frequency, appearing 3,021 times, and the word *apricot* has low frequency, appearing 286 times, among the 131 million words in the Hyperspace Analogue to Language corpus (Balota et al., 2007; Lund & Burgess, 1996). The benefit of low-frequency (LF) words over high-frequency (HF) words in single-item recognition is well documented and considered a regularity of memory (Glanzer, Adams, Iverson, & Kim, 1993). Typically this word frequency effect manifests as a mirror pattern, where hit rates (HRs) are higher and false alarm rates (FARs) are lower for LF words than for HF words (e.g., Glanzer & Adams, 1985). This WF mirror effect is a benchmark finding that is accounted for by most quantitative models of recognition memory, albeit with different underlying mechanisms. For example, some models assume that the uncommon features of LF words make them particularly diagnostic, thereby increasing the HR (e.g., Shiffrin & Steyvers, 1997). Common features, on the other hand, make HF words more similar to other words and therefore more confusable, increasing the FAR. Other models assume that the large number of preexperimental contexts in which HF words are experienced creates confusion and reduces accuracy (e.g., Dennis & Humphreys, 2001; Reder et al., 2000). All successful models of recognition memory, regardless of mechanism, predict a WF mirror pattern.

One mainstay of studies evaluating WF is that WF is routinely treated as binary, with low and high categories (cf. Criss & Malmberg, 2008). In the natural world, however, WF spans a very wide range. What remains unknown is whether the categorization of WF into two discrete categories obscures the true underlying relationship between frequency and memory. While discretizing a large continuous variable provides a powerful tool for data visualization and analysis, this might also lead to mischaracterizations of the underlying relationship between the stimulus and the observed behavior. Estes (e.g., 1956) long cautioned against the uncritical use of averaged participant effects, and the same is true at the stimulus level (e.g., H. H. Clark, 1973; Freeman, Heathcote, Chalmers, & Hockley, 2010).

Evaluations of the effect of a third category of words, very low frequency (VLF) words, offer tantalizing evidence that the relationship between WF and recognition memory might in fact be more complicated. Schulman (1976) first showed that recognition performance is lower for rare words rated as nonwords than for VLF words rated as familiar. A number of subsequent studies replicated the finding of lower accuracy for VLF than LF words (Mandler, Goodman, & Wilkes-Gibbs, 1982; Wixted, 1992; Zechmeister, Curt, & Sebastian, 1978). Rao and Proctor (1984) replicated this inverted U-shaped function and also found that the pattern endures under self-paced study. However, interpretation of these findings is complicated by the fact that the VLF words were typically judged nonwords by participants. Further, none of the

Pernille Hemmer, Department of Psychology, Rutgers University; Amy H. Criss, Department of Psychology, Syracuse University.

This work was made possible by National Science Foundation Grant 0951612 awarded to Amy H. Criss. These data were previously presented at the annual meeting of the Society for Mathematical Psychology and at the annual meeting of the Psychonomic Society.

Correspondence concerning this article should be addressed to Pernille Hemmer, Department of Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854. E-mail: pernille.hemmer@rutgers.edu

studies evaluated WF as the continuous variable that it is. In fact, Schulman (1976) noted that "no one has bothered to test recognition memory over the whole range of word frequency" (p. 301)—a problem that persists today.

The goal of this article is to characterize the relationship between the continuous property of normative WF and single-item recognition performance. The resulting data have important implications for both existing and future theories of episodic memory. We developed a large stimulus set that contains word frequencies from a broad range. Each participant saw a subset of words randomly sampled across the frequency range. In contrast to previous work, we analyzed performance across the WF range, both for individual WF values and across participants.

Method

Participants

Four hundred sixty-two undergraduate students at Syracuse University participated in exchange for course credit. Seventy-two participants did not finish all 15 blocks due to time constraints or computer malfunction. Of those, 49 completed all three blocks of single-item recognition, 21 completed two blocks, and two completed one block. We repeated all analyses on the subset of 390 participants who completed all 15 blocks, and there were no differences in the results from the full set of 462 participants reported in this article.

Materials

We developed a stimulus set of 924 words extracted from the Touchstone Applied Science Associates corpus (Landauer, Foltz, & Laham, 1998). WF values were selected to represent a broad range of values (283–1,358 words per million) and to reduce the correlation between WF and context variability. For each of the 924 words in the data set, we obtained the WF values from Kucera and Francis (1967) for archival comparison to the literature in which Kucera and Francis is the standard value reported. Our stimulus set ranged from 1 to 197 words per million in Kucera and Francis WF values.

Design and Procedure

The experiment included five different tasks: single-item recognition, associative recognition, cued recall, free recall, and lexical decision (Hemmer & Criss, 2012a, 2012b). Each task was repeated three times over the course of the experiment for a total of 15 blocks. The first five blocks consisted of the first presentation of each of the five tasks (randomly ordered for each participant). For the remaining 10 blocks the five task types were presented twice in random order. Task was postcued; therefore participants could not adopt a study strategy based on the anticipated test type.

All blocks (except lexical decision) began with a study phase in which participants viewed 20 word pairs presented side by side, one pair at a time. Each pair remained on the screen for 2 s and was immediately followed by text asking participants to "Please rate the degree of association between the two items you just saw" on a scale ranging from 1 (*not at all associated*) to 9 (*highly associated*). The word pair was not visible on the screen during the rating. Responses were self-paced by clicking on boxes numbered 1–9 on the screen.

Each study phase was immediately followed by a distractor task. This was a simple math task where participants continuously added a series of 15 random digits drawn with replacement from the range 1–9. Digits were presented at a rate of 3 s per digit, for a total presentation time of 45 s. After all digits appeared, participants typed in their response and received accuracy feedback.

For the remainder of the article we discuss the procedures and results related to only the single-item recognition task. For the test stimuli, 10 study items were selected at random from the study list. The 10 items could be from either the right or the left presentation position but not from both the left and the right presentation position for the same study trial. These 10 old items were combined with 10 foil words that had not been previously studied. The test words were randomly ordered and presented one at a time in the center of the screen. Participants are asked to "indicate if the item you see on the screen was on the list you just studied (YES) or not on the list (NO)." Participants responded by clicking on boxes presented on the computer screen. Responses were self-paced.

Each study-distractor-test block was followed by the option to take a self-paced break. The experiment lasted approximately 1 hr.

Results

Because the distribution of Kucera and Francis (1967) WF values in the data set were heavily right-skewed, all WF values were transformed to log scale prior to statistical analysis to meet the assumption of normality.

Standard WF Analysis

We first analyzed the data in the standard categorical fashion of assuming WF takes either high or low values. As there are no well-defined values that constitute high and low, we chose high and low groupings based on publications by one of the authors (e.g., Criss & McClelland, 2006; Criss & Shiffrin, 2004). We chose LF values as those between 1 and 10 (corresponds to values of less than 1 on a log scale) and HF values as those greater than 50 (greater than 1.7 on a log scale). Figure 1 shows the ubiquitous mirror effect with a low frequency advantage for hits, confirmed by a paired t test, $t(267) = 3.74$, $p < .001$, and for false alarms, $t(278) = -4.94$, $p < .001$ (any participant who did not have data in both the LF and HF bin was dropped from the analysis). We completed this analysis for a large number of other combinations of high and low values from the literature. The results always presented as a low-frequency advantage and a mirror effect.

Item-Level Analysis

We conducted an item-level analysis to assess the relationship between continuous WF and recognition performance at the individual WF value. The stimulus set contained 133 unique WF values, and we computed the average HR and FAR for each value and then used those as predictor values in the analysis. Figure 2 shows the relationship between WF and recognition performance across the 133 unique WF values.

To characterize the observed nonmonotonic U-shaped relationship between HR and WF, two regression models were fit to the

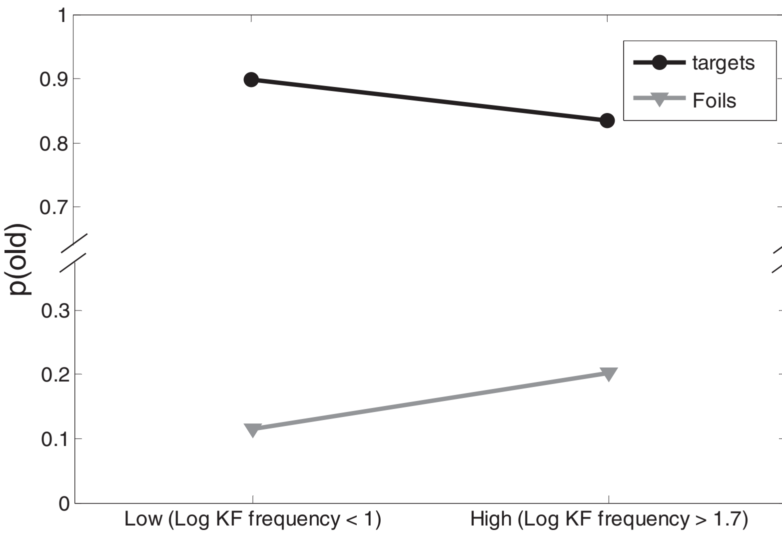

Figure 1. A traditional word-frequency analysis showing hit rates (circles) and false alarm rates (triangles) as a function of categorical word frequency. Word frequency is given in log-scaled Kucera and Francis (KF) frequency values.

data. The first model was a linear model intended to test whether the linear relationship predicted in the literature based on binary evaluations of WF holds for continuous levels of WF used here. The second model was a quadratic model intended to test whether a nonlinear relationship provides a better description of continuous WF. Both models were evaluated for goodness-of-fit compared to the constant model (intercept: 0.848) and for improvement in R^2 over the previous model. The linear model Y = 0.873 − 0.015X did not provide a significantly better fit than the constant model (R^2 = .007), *F*(1,

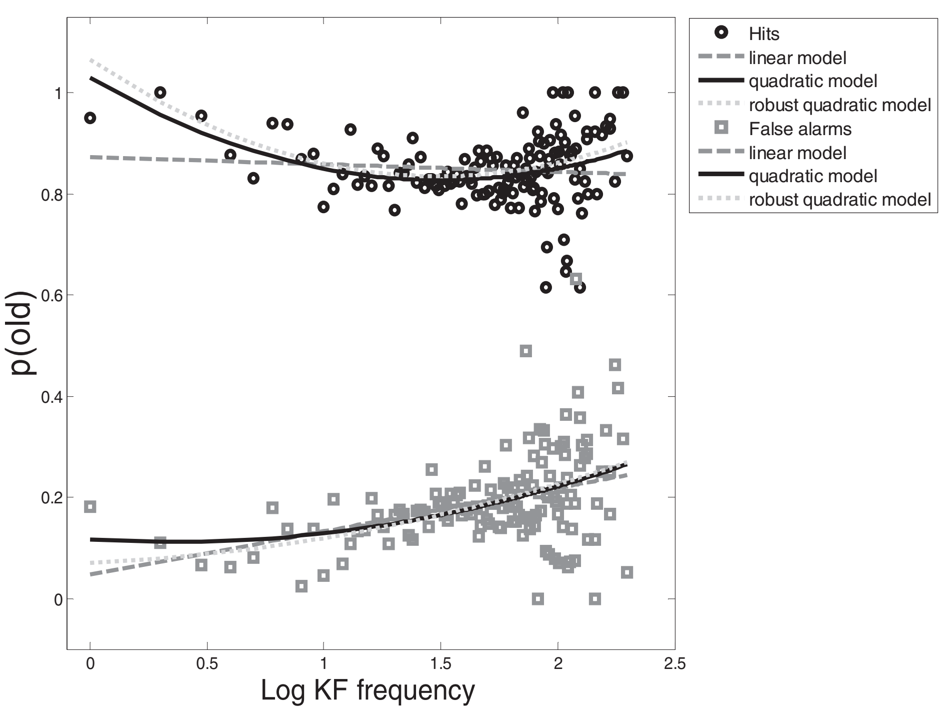

Figure 2. Item-level analysis showing hit and false alarm rates as a function of word frequency for each unique word frequency value. Black circles and gray squares represent word frequency values averaged over words. The dashed lines give the linear fit, the solid lines the quadratic fit, and the dotted lines the robust quadratic fit. Word frequency is given in log-scaled Kucera and Francis (KF) frequency values.

131) = 0.98, p = .323. The quadratic model Y = 1.029 − 0.269X + 0.090X^2 did provide a significantly better fit (R^2 = .12), $F(2, 130) = 8.69$, $p < .001$, and it was significantly better than the linear model, $F(2, 130) = 16.36$, $p < .001$.

Two regression models were also fit to the FAR at the item level. The linear model Y = 0.047 + 0.086X provided a significantly better fit than the constant model (intercept: 0.194, R^2 = .15), $F(1, 131) = 23.1$, $p < .001$. The quadratic model Y = 0.117 − 0.028X + 0.040X^2 also provided a significantly better fit (R^2 = .16), $F(2, 130) = 12.7$, $p < .001$. However, the quadratic model did not provide any further reduction in R^2 compared to the linear model. Thus, the false-alarm data shows the characteristic relationship of lower FAR for LF than for HF words across the full range of WF.

A robust quadratic model was fit to the data to discount the possible influence of outliers. For HR the robust model fit better than the linear model, $F(2, 130) = 23$, $p < .001$, and reduced R^2 = .26 over the quadratic model, $F(2, 130) = 25.16$, $p < .001$. For FAR, the robust model fit better than the linear model, $F(2, 130)$ = 18, $p < .001$, and reduced R^2 = .15 over the quadratic model, $F(2, 130) = 8.97$, $p < .001$. Observation of Figure 2 indicates that the quadratic fits were not substantially affected by outliers. Note that Figure 2 shows much higher variability in performance for both HRs and FARs for very high frequency words. However, this is likely due to measurement noise rather than being indicative of a true property of the cognitive system. As mentioned, there were 924 words but only 133 unique WF values. The number of unique words per value ranged from 1 to 25 (M = 6.95), and very high frequency values (i.e., log WF ≥ 2) tended to have many fewer unique words per value (M = 1.44).

Participant-Level Analysis

For statistical analysis, the WF values were binned into 14 approximately equal sized bins (because 924 —the number of stimuli—is evenly divisible by 14). Because many words in the stimulus set share the same WF value, it was not possible to partition the stimulus such that exactly 66 items fell in each bin. However, each bin contained between 6.1% and 8% of the total word set. Accuracy was calculated for each bin, for each block, and for each participant and then averaged over blocks for each participant and finally averaged for participants for each bin.

The relationship between WF and memory is similar when the data are analyzed across participants (see Figure 3) and when the data are analyzed across items (see Figure 2). HRs for both LF and HF words were better than recognition for the midrange frequency words. To confirm the nonmonotonicity of the relationship, two regression models were fit to the data using the same statistical approach as for the item-level analysis. The linear model Y = 0.828 − 0.014X did not provide a significantly better fit than the constant model (intercept: 0.828, R^2 = .04), $F(1, 12) = 0.49$, p = .497. The quadratic model Y = 1.120 − 0.433X + 0.144X^2,

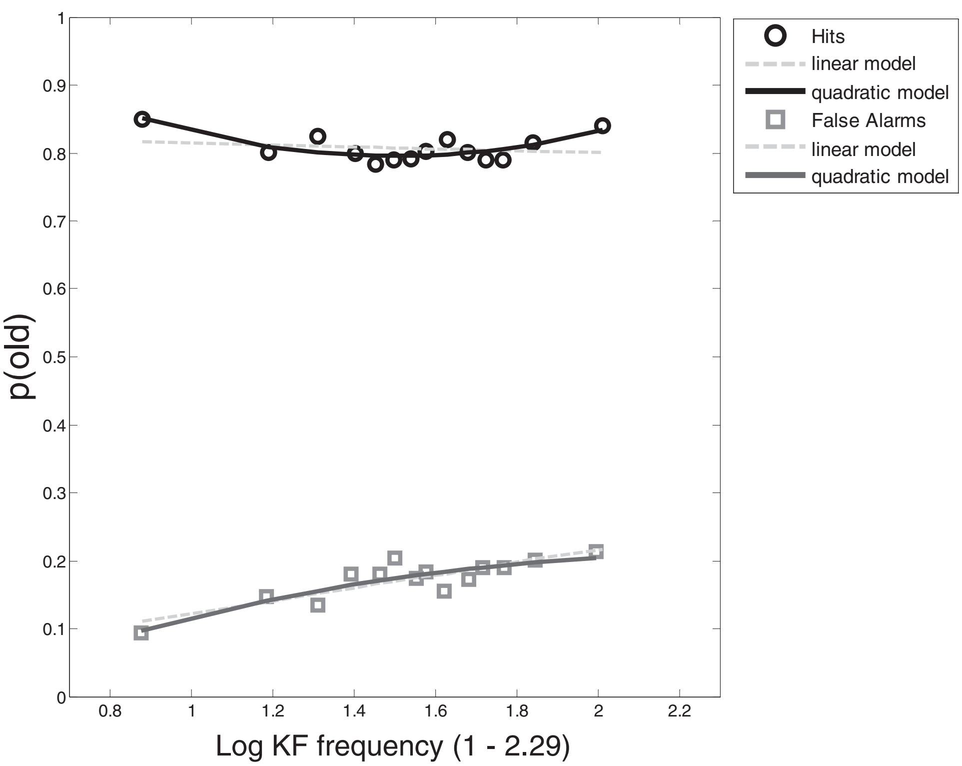

Figure 3. Participant-level analysis showing hit rate (circles) and false alarm rate (squares) as a function of word frequency binned and averaged over participants. The word frequency value is the average value for each bin. Dashed lines give the linear fit, solid lines the quadratic fit. Word frequency is given in log-scaled Kucera and Francis (KF) frequency values.

however, provided a significantly better fit (R^2 = .65), $F(2, 11) = 10.4$, $p = .003$, and it was significantly better than the linear model, $F(2, 11) = 19.46$, $p < .001$. Two regression models were also fit to the FAR data at the participant level. The linear model $Y = 0.0276 + 0.095X$ provided a significantly better fit than the constant model (intercept: 0.096, $R^2 = .73$), $F(1, 12) = 33$, $p < .001$, as did the quadratic model $Y = -0.084 + 0.256X - 0.056X^2$ ($R^2 = .77$), $F(1, 11) = 18.4$, $p < .001$. The quadratic model, however, did not provide a further reduction in R^2, $F(2, 11) = 1.71$, $p = .226$. Thus, the participant-level analysis replicates the item-level analysis in all respects.

Discussion

Our analysis of WF as a continuous variable in single-item recognition shows a clear nonmonotonic U-shaped pattern for hit rates. Both low-frequency and high-frequency words have higher hit rates than do midrange-frequency words when analyses are conducted at the item and the participant level. The U-shaped HR function is complemented by a linear FAR function across WF. By evaluating WF beyond the standard high/low categories, we have uncovered a more complex relationship than was previously known.

It is possible that other properties, such as context variability or word length, which may be correlated with WF, play a role in the observed pattern. The field has adopted WF as the theoretically important property and the focus of empirical and theoretical development, and we follow suit. However, we explicitly chose our stimulus set to minimize the correlation between WF and context variability. While this does not rule out the role of correlated variables, we are reasonably confident that WF itself is the critical variable here. We also acknowledge that it is perhaps somewhat unusual that we included multiple memory tasks for each participant or that participants studied pairs. However, many published articles have found standard WF effects under such circumstances (e.g., S. E. Clark, 1992; S. E. Clark & Shiffrin, 1992; Dorfman & Glanzer, 1988; Gillund & Shiffrin, 1984; Hockley, 1992, 1994; May, Cuddy, & Norton, 1979). Further, we too demonstrate a typical WF mirror effect when we follow tradition and treat WF as a categorical variable with the values of high and low (see Figure 1). Therefore, there is no reason to believe that the specific procedures used here lead to the observed nonlinearity in the data.

Our pattern of data presents a problem for all models of recognition, which, by design, predict higher HRs and lower FARs for LF words. Explanations of the WF effect are many, but we consider three broad categories. First, LF words are better remembered because they attract more attention during encoding (e.g., Criss & Malmberg, 2008; Glanzer & Adams, 1990; Malmberg & Nelson, 2003). Second, LF words are better remembered because they have fewer preexperimental contexts causing less interference with memory for the experimental context (Dennis & Humphreys, 2001; Reder et al., 2000). Third, LF words are composed of less common and therefore more distinctive diagnostic features, leading to better memory (Landauer & Streeter, 1973; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). Any of these classes of models could likely be modified to account for the data wherein HRs have a nonmonotonic and FARs have a linear relationship with WF. However, the modifications are not immediately obvious. Further, it seems necessary to posit a second mechanism to account for the data. That is, it seems implausible to assume that both lower and very high frequency words receive more attention or that both very few and very many preexperimental contexts contribute less interference.

One explanation that could be added to any of the above models to account for this data is based on expectations developed from prior experiences. Suppose that, as predicted by most models, memory accuracy is higher for lower frequency words due to diagnostic item features, better binding to the experimental context, attention, or some other mechanism. Further, suppose that when memory is weak, participants make educated guesses about whether a word was studied. That is, the guess is based on knowledge developed over the course of life prior to the current situation (the experiment, in this case). Very high frequency words are encountered frequently (by definition), and thus an educated guess would predict that a very high frequency word was in fact studied (regardless of the true status of the test item), elevating both hits and false alarms for very HF words. We suggest that the data presented here might be explained by a model that combines higher accuracy for lower frequency words and a bias to claim that very high frequency words were studied based on prior experience. Such a mechanism is not present in any existing model of memory but could potentially be added with appropriate modifications.

In summary, the relatively simple but elegant methods of measuring WF as the continuous variable that it is, combined with analyzing data at both the item and participant levels, has revealed a new finding: a U-shaped HR function that is inconsistent with the hallmark WF mirror effect. Model development must incorporate this new finding, and we propose considering the role that expectations based on prior knowledge play in making episodic memory judgments.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. doi:10.3758/BF03193014

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359. doi:10.1016/S0022-5371(73)80014-3

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20,* 231–243. doi:10.3758/BF03199660

Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition, 20,* 580–598. doi:10.3758/BF03199590

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

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language, 55,* 447–460. doi:10.1016/j.jml.2006.06.003

Criss, A. H., & Shiffrin, R. M. (2004). Interactions between study task, study time, and the low frequency hit rate advantage in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 778–786. doi:10.1037/0278-7393.30.4.778

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. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53,* 134–140. doi:10.1037/h0045156

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language, 62,* 1–18. doi:10.1016/j.jml.2009.09.004

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. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16. doi:10.1037/0278-7393.16.1.5

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

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 21–31. doi:10.1037/0278-7393.2.1.21

Hemmer, P., & Criss, A. H. (2012a, July). *The effect of continuous word frequency, context variability and OLD across memory tasks*. Poster presented at the annual meeting of the Society for Mathematical Psychology, Columbus, Ohio.

Hemmer, P., & Criss, A. H. (2012b, November). *The effect of continuous word frequency, context variability and OLD across memory tasks*. Poster presented at the annual meeting of the Psychonomic Society, Minneapolis, Minnesota.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1321–1330. doi:10.1037/0278-7393.18.6.1321

Hockley, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition, 22,* 713–722. doi:10.3758/BF03209256

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25,* 259–284. doi:10.1080/01638539809545028

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12,* 119–131. doi:10.1016/S0022-5371(73)80001-5

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers, 28,* 203–208. doi:10.3758/BF03204766

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

Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition, 10,* 33–42. doi:10.3758/BF03197623

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

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 724–760. doi:10.1037/0033-295X.105.4.734-760

Rao, K., & Proctor, R. W. (1984). Study-phase processing and the word frequency effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 386–394. doi:10.1037/0278-7393.10.3.386

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (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

Schulman, A. (1967). Word length and rarity in recognition memory. *Psychonomic Science, 9,* 211–212.

Schulman, A. (1976). Memory for rare words previously rated for familiarity. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 301–307. doi:10.1037/0278-7393.2.3.301

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

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 681–690. doi:10.1037/0278-7393.18.4.681

Zechmeister, E. B., Curt, C., & Sebastian, J. A. (1978). Errors in a recognition memory task are a U-shaped function of word frequency. *Bulletin of the Psychonomic Society, 11,* 371–373.

Received March 13, 2013
Revision received June 3, 2013
Accepted June 4, 2013 ■