

# Feature frequency effects in recognition memory

KENNETH J. MALMBERG, MARK STEYVERS, JOSEPH D. STEPHENS, and RICHARD M. SHIFFRIN  
*Indiana University, Bloomington, Indiana*

Rare words are usually better recognized than common words, a finding in recognition memory known as the *word-frequency effect*. Some theories predict the word-frequency effect because they assume that rare words consist of more distinctive features than do common words (e.g., Shiffrin & Steyvers's, 1997, REM theory). In this study, recognition memory was tested for words that vary in the commonness of their orthographic features, and we found that recognition was best for words made up of primarily rare letters. In addition, a mirror effect was observed: Words with rare letters had a higher hit rate and a lower false-alarm rate than did words with common letters. We also found that normative word frequency affects recognition independently of letter frequency. Therefore, the distinctiveness of a word's orthographic features is one, but not the only, factor necessary to explain the word-frequency effect.

Rare words are better recognized than common words (Schulman, 1967; Shepard, 1967; but see Wixted, 1992), a finding in recognition memory known as the *word-frequency effect* (WFE). For single-word old-new recognition, hit rates (HRs, correctly responding "old" to an old word) are higher and false-alarm rates (FARs, incorrectly responding "old" to a new word) are lower for low-frequency (LF) words than for high-frequency (HF) words (Glanzer & Adams, 1985). Examples of some of the accounts for the advantage for LF words include differences in the distribution of attentional resources (e.g., Glanzer & Adams, 1990; Maddox & Estes, 1997), multiple retrieval processes (e.g., Joordens & Hockley, 2000), the number of different contexts in which words appear (e.g., Dennis & Humphreys, 2001), and differences in encoding variability (e.g., McClelland & Chappell, 1998). No consensus yet exists regarding how many—if any—of these accounts is correct, but it is quite clear that normative word frequency is correlated with a large number of variables that could theoretically produce the WFE (cf. Gillund & Shiffrin, 1984; Shiffrin & Steyvers, 1997). In the present study, we directly tested the hypothesis (not listed above) that differences in the distinctiveness of the features that comprise words of differing frequency produce the WFE (e.g., Malmberg & Murnane, 2002; Shiffrin & Steyvers, 1997).

One factor that is correlated with normative word frequency is normative letter frequency (Landauer & Streeter, 1973), a fact that is consistent with the hypothesis that LF words are better recognized than HF words because the memory representations of LF words are more "distinctive" than memory representations of HF words. Conver-

gent evidence for the distinctiveness hypothesis comes from a set of experiments by Zechmeister (1969, 1972; also see Hunt & Elliott, 1980), who showed that words rated by subjects as being orthographically distinct (e.g., *sylyph*) were better recognized than words rated as less orthographically distinct (e.g., *parse*). A shortcoming of these studies is that a concrete definition of what it means to be "distinctive" has been difficult to achieve. For example, it is not at all surprising to find a distinctiveness effect if subjects have some sort of bias or strategy to rate LF words as being relatively distinct. In the present study, we examined the feature-frequency assumption of *retrieving effectively from memory* (REM; Shiffrin & Steyvers, 1997, 1998) by varying the orthographic-feature frequency of studied and tested words. First, however, we discuss in detail the REM account of the WFE.

## REM

The REM theory offers a concrete example of the feature-distinctiveness account of the WFE (Shiffrin & Steyvers, 1997, 1998). REM's Bayesian framework assumes that features vary in their environmental frequency, or base rate, and rare features are relatively more "diagnostic" in REM. A match between a rare probe feature and a corresponding feature in memory provides more evidence in favor of the probe being "old" because rare features are unlikely to be encountered by chance alone (see the Bayesian calculations given below). Thus, REM accounts for the WFE by assuming that the memory representations of LF words tend to be made up of less common and therefore more diagnostic features than do the memory representations of HF words (Malmberg & Murnane, 2002). We term this the *feature-frequency* assumption. For this reason, REM predicts a "mirror-patterned" (Glanzer & Adams, 1985) advantage for LF words over HF words: For yes-no recognition, the probability of responding "old" to an old item (i.e., HR) is greater for LF words than for HF words, and the probability of responding "old" to a new test item (i.e., FAR) is less for LF words than for HF words.

---

NIMH Grant 12643 and an NIMH Mathematical Modeling Training Grant to Indiana University supported the work of K.J.M. NIMH Grants 12717 and 63993 supported the work of R.M.S. K.J.M. thanks Eric-Jan Wagenmakers for providing the motivation needed to complete this research. Correspondence should be addressed to K. J. Malmberg, Department of Psychology, Indiana University, Bloomington, IN 47405 (e-mail: malmberg@indiana.edu).

Specifically, REM assumes that separate memory traces (*images*) represent different items, and a vector,  $\mathbf{V}$ , of  $w$  features, comprises a trace. Generic knowledge is stored in lexical/semantic images, and every known word has a lexical/semantic image consisting of  $w$  features each greater than zero<sup>1</sup> (0 is used to represent no knowledge concerning a feature, and episodic vectors will often contain such indicators). The probability of observing feature value,  $j$ , is governed by a geometric probability distribution:

$$P(V = j) = (1 - g)^{j-1}g, \quad j = 1, \dots, \infty, \quad (1)$$

where  $g$  determines the frequency and the variability of different features in the environment. When  $g$  is relatively high, the features drawn from the distribution will tend to be integers with relatively small values. When  $g$  is relatively low, values drawn from the distribution are more varied with a greater mean. The assumption that HF words have more common feature values has been instantiated in REM by assuming that HF words have higher values of  $g$ .

When a word is studied, an episodic image of its lexical semantic image is stored in memory, and images of different words are stored in different vectors. After  $t$  time units of study, the probability that a feature will be stored in the episodic image is  $1 - (1 - u^*)^t$ ; otherwise, 0 is stored ( $u^*$  is the probability of storing a feature in a unit of time). If storage of a value occurs, the feature value is correctly copied with probability  $c$ . With probability  $1 - c$  the value stored is sampled randomly according to Equation 1. Note that random sampling means that a value can be stored correctly by chance and, more important, that this occurs more often for common values; it is this fact that underlies the REM account of the WFE.

At test, the  $w$  item features of the lexical/semantic vector representing the test item serve as a retrieval cue. The cue is matched with the  $n$  episodic images ( $I_j$ ) in memory, and the system notes which features of  $I_j$  match the corresponding feature of the cue and the matching value ( $n_{ijm}$  stands for the number of matching values in the  $j$ th image that have value  $i$ ) and which features mismatch ( $n_{jq}$  stands for the number of mismatching values in the  $j$ th image). Next, a likelihood ratio,  $\lambda_j$ , is computed for each  $I_j$ :

$$\lambda_j = (1 - c)^{n_{jq}} \prod_{i=1}^{\infty} \left[ \frac{c + (1 - c)g(1 - g)^{i-1}}{g(1 - g)^{i-1}} \right]^{n_{ijm}}, \quad (2)$$

where  $g$  is the long-run environmental base rate for the occurrence of features.  $\lambda_j$  is the likelihood ratio for the  $j$ th image and can be thought of as a match strength between the retrieval cue and  $I_j$ .

The recognition decision is based on the odds,  $\Phi$ , the probability that the test item is old divided by the probability the test item is new (Shiffrin & Steyvers, 1997):

$$\Phi = \frac{1}{n} \sum_{j=1}^n \lambda_j, \quad (3)$$

where  $n$  is the number of items studied. If the odds exceed a criterion, then an “old” response is made. The default criterion is 1.0.

REM predicts an LF HR advantage because it assumes that LF words consist of more uncommon features than HF words (i.e.,  $g_{HF} > g_{LF}$ ), and matching uncommon features contributes more to  $\lambda_j$  than matching common features according to REM’s Bayesian calculation of activation (Equation 2). Figure 1 shows that the LF advantage increases as the difference between  $g_{HF}$  and  $g_{LF}$  increases. That is, LF words are better recognized relative to HF words as the features that make up LF words become more distinct (cf. Malmberg & Murnane, 2002).

At an intuitive level, Equation 2 determines how likely it is that a retrieval cue and an image in memory represent the same word given the set of features they have that share the same value and the features they have that differ in value. If the cue and an image are different, then features are expected to match only by chance. Because uncommon features are less likely to match by chance, they provide more “diagnostic” matching information. Thus, LF targets produce relatively greater amounts of familiarity at test because their features produce greater levels of activation than do HF targets (on average). For foils, however, every feature match is spurious, and this occurs relatively infrequently for LF words. Thus, the FAR effect is predicted because the common features making up HF retrieval cues tend to match the images of other words in memory more often by chance than do the uncommon features making up LF retrieval cues. When familiarity is calculated according to Equation 3, the normalized sum of HF  $\lambda_j$ s tends to be greater than the normalized sum of LF  $\lambda_j$ s.

Note that diagnostic features are unusual features, and, therefore, the LF words they tend to make up are relatively distinct. HF words tend to be less distinct because they tend to share relatively common features, and, therefore, they are more similar than LF words. Thus, the REM concept of “diagnosticity” is closely related to the “distinctiveness” account of the WFE, because rare features are more distinctive than common features.

## EXPERIMENT

Here, we directly tested the hypothesis that the frequency of occurrence of orthographic features (or letters) in natural language affects the recognition of words. According to the feature-frequency account of REM (Shiffrin & Steyvers, 1997; also see Zechmeister, 1969, 1972), words composed primarily of LF letters should be better recognized than words composed primarily of HF letters. Alternatively, if orthographic-feature frequency does not affect word recognition, then words composed of common letters (as opposed to uncommon letters) should be recognized equally well.

In addition to orthographic-feature frequency, normative word frequency was manipulated in this experiment. If orthographic-feature frequency accounts for the entire WFE, then one would expect to see LF words and HF words recognized equally well when orthographic-feature frequency is held constant. However, this would be highly unlikely since we were manipulating only a subset of features that represent words (e.g., leaving out semantic features);

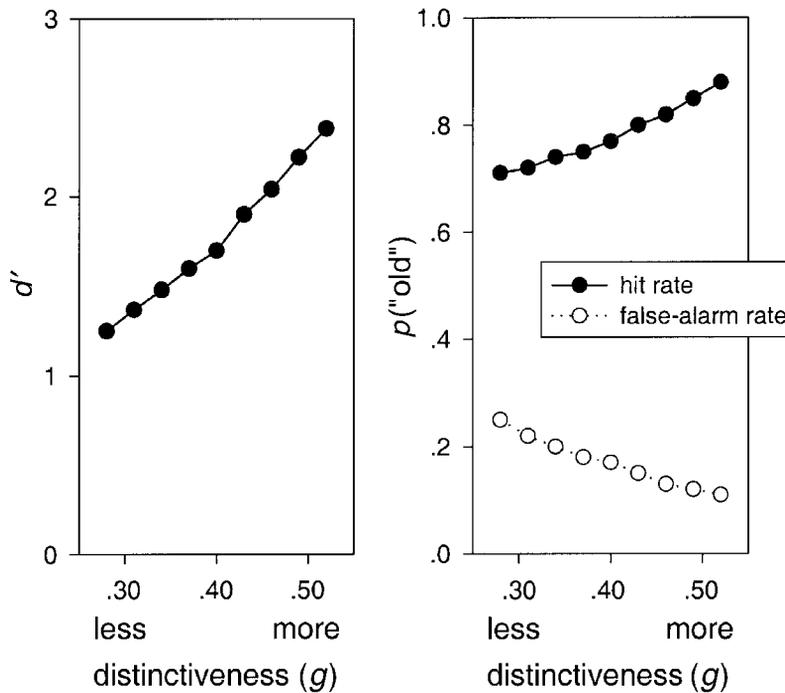


Figure 1. Recognition performance as a function of feature distinctiveness ( $g$ ) in REM.

therefore, we expected both orthographic-feature frequency and normative word frequency to have significant effects on recognition.

## Method

**Subjects.** Fifty-three Indiana University students who were enrolled in introductory psychology courses participated in exchange for course credit.

**Design and Materials.** Normative word frequency and normative letter frequency were manipulated as within-subjects factors in a  $2 \times 2$  factorial design. The dependent variables were HRs, FARs, and  $d_a$  (Macmillan & Creelman, 1991).

Two hundred eighty-eight words were selected from Kučera and Francis (1967). Each word was between four and seven letters long (inclusive). These words are listed in Appendix A. The number of times a letter occurred in the corpus in each position was multiplied by the sum of the normative frequency of the words that contained a letter in a given position. We used three different counts corresponding to the first, interior, and final letter positions of a word because different letters occur in different positions with different relative frequencies. Thus, if a letter occurred in a given position only once, and the word in which it appeared normatively occurred  $n$  times per million, then the weighted letter count was  $1n$ . The weighted letter counts were then normalized so that the sum of the 26 individual letter counts equaled 1 by dividing each weighted letter count by the sum of the weighted letter counts. In other words, we computed the relative frequency with which a letter was normatively expected to be encountered in each of the three different word positions. Table 1 lists these relative orthographic-feature frequencies.<sup>2</sup>

The distinctiveness of a given word was determined by computing the average orthographic-feature frequency of letters of which it was composed (referred to as *mean letter-frequency*). Thus, words that consisted of relatively LF letters produced relatively low mean letter frequencies. For example, consider the words BANE and AJAR as

examples of LF words with letters that differ in their mean letter frequencies. Table 1 lists the relative frequencies of occurrence for each letter for the first, interior, and final positions of a word. The words BANE and AJAR get mean letter-frequency scores of  $(.0562 + .1071 + .0622 + .2157) / 4 = .11$  and  $(.0498 + .00007 + .1071 + .0975) / 4 = .06$ , respectively. Thus, the word BANE tends to consist of more HF letters than the word AJAR.

The stimuli were organized into four groups of 72 by crossing orthographic-feature frequency and normative word frequency. The conditions simultaneously satisfied three constraints: (1) HF words and LF words were operationally defined as those occurring between 15 and 39 times and between 3 and 7 times per million words, respectively, (2) the mean letter frequencies for the HF and LF words were equated as nearly as possible, and (3) each condition had approximately equal numbers of four-, five-, six-, and seven-letter words. Appendix B lists the mean normative word frequencies and the mean letter frequencies for the four word groups.

Each study list consisted of 130 words: 24 words from each of the four conditions and 34 filler items. Study position was randomly determined for each critical word for each subject, except for the primacy and recency buffer words, which were always filler items. Twelve targets and 12 distractors were randomly selected from each condition and were randomly assigned a position on the 96-item test lists.

**Procedure.** An experimental session consisted of two study-test cycles. The subjects were instructed prior to each study-test cycle to remember the words on the study list for a later unspecified memory test. Each word was displayed in uppercase in the center of the computer screen for 1.3 sec of study.

At test, subjects performed a series of single-item confidence ratings trials. Test items were presented one at a time, and the subjects were instructed to rate how confident they were that a test item was studied by utilizing a 6-point scale (a 1 indicated the lowest confidence that a word had been studied, and a 6 indicated highest confidence that an item had been studied). The subjects responded by using a pointing device (mouse) to select the appropriate button in the com-

**Table 1**  
**Relative Frequencies of Letters in First, Interior, and Last Word Positions**

| Rank |   | First     |   | Interior  |   | Last      |
|------|---|-----------|---|-----------|---|-----------|
| 1    | t | 0.1396    | e | 0.1248    | e | 0.2157    |
| 2    | w | 0.1103    | a | 0.1071    | t | 0.1463    |
| 3    | s | 0.0997    | i | 0.1015    | d | 0.0875    |
| 4    | f | 0.0592    | o | 0.0969    | r | 0.0837    |
| 5    | m | 0.0587    | r | 0.0790    | n | 0.0824    |
| 6    | c | 0.0587    | h | 0.0672    | y | 0.0747    |
| 7    | b | 0.0562    | n | 0.0622    | h | 0.0672    |
| 8    | a | 0.0498    | t | 0.0560    | l | 0.0538    |
| 9    | h | 0.0481    | l | 0.0548    | s | 0.0428    |
| 10   | p | 0.0462    | u | 0.0494    | g | 0.0406    |
| 11   | l | 0.0428    | s | 0.0394    | m | 0.0324    |
| 12   | d | 0.0328    | c | 0.0335    | k | 0.0234    |
| 13   | r | 0.0318    | m | 0.0209    | o | 0.0108    |
| 14   | e | 0.0283    | v | 0.0194    | w | 9.4590e-3 |
| 15   | o | 0.0255    | g | 0.0176    | p | 7.9290e-3 |
| 16   | g | 0.0250    | d | 0.0156    | f | 6.9520e-3 |
| 17   | n | 0.0193    | p | 0.0141    | a | 5.8730e-3 |
| 18   | i | 0.0165    | w | 8.0830e-3 | c | 5.3200e-3 |
| 19   | u | 0.0136    | f | 8.0480e-3 | b | 1.1920e-3 |
| 20   | v | 0.0107    | b | 7.9900e-3 | x | 9.2800e-4 |
| 21   | j | 8.6410e-3 | k | 7.2420e-3 | l | 3.6300e-4 |
| 22   | k | 8.2810e-3 | y | 3.8800e-3 | z | 3.5800e-4 |
| 23   | y | 7.5650e-3 | x | 2.8410e-3 | u | 2.6700e-4 |
| 24   | q | 2.5860e-3 | z | 9.2900e-4 | j | 0.0000    |
| 25   | z | 2.3800e-4 | q | 8.1800e-4 | q | 0.0000    |
| 26   | x | 1.7000e-5 | j | 7.0000e-4 | v | 0.0000    |

Note—Letter counts were weighted with the Kučera and Francis (1967) frequency counts of the words they appeared in.

puter display. Each response was followed immediately by the presentation of the next test item.

## Results

An alpha of .05 was the standard of significance for all statistical analyses. The 6-point confidence ratings were used to compute 5-point  $z$ -transformed ratings ROC curves for each condition and subject. The slopes of the individual  $z$ -transformed ROCs were then used to compute  $d_a$  (Macmillan & Creelman, 1991). To obtain HRs and FARs, the confidence ratings were converted to binary “old”–“new” responses by classifying ratings greater than or equal to a criterion as “old” responses. For each subject, the criterion was chosen to equalize the overall number of “old” and “new” responses as best as possible.<sup>3</sup>

The top panel of Figure 2 show plots  $d_a$  as a function of normative word frequency and orthographic-feature frequency. It shows that words consisting primarily of LF letters were better recognized than words consisting primarily of HF letters. The mean  $d_a$  was greater for LF letters than for HF letters [ $F(1,52) = 103.2, MS_e = 0.13$ ]. In the bottom panel of Figure 2, the mean probability of responding “old” is shown for the targets and distractors for these four conditions. The HR for words with LF letters was slightly greater than the HR for words with HF letters features [ $F(1,52) = 2.56, MS_e = 0.01, p \leq .12$ ], and the FARs were significantly lower for words with LF letters than for words with HF letters [ $F(1,52) = 31.10, MS_e = 0.01$ ].

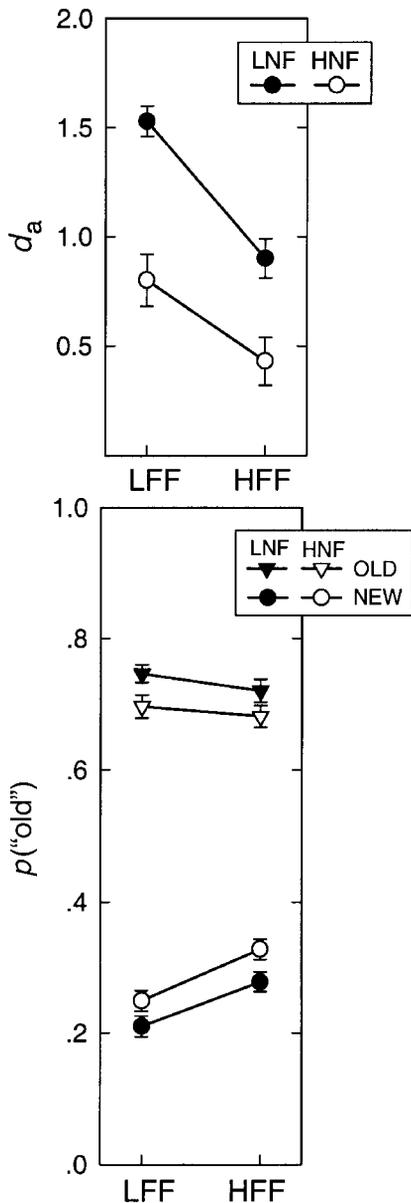
Figure 2 also illustrates that LF words were better recognized than HF words. The mean  $d_a$  was greater for LF words than for HF words [ $F(1,52) = 45.78, MS_e = 0.42$ ]. The bottom panel of Figure 2 shows that a typical mirror-

patterned WFE was observed: HRs were significantly higher for LF words than for HF words [ $F(1,52) = 11.77, MS_e = 0.01$ ], and FARs were significantly lower for LF than for HF words [ $F(1,52) = 11.65, MS_e = 0.01$ ]. The interaction between word and letter frequency factors was significant [ $F(1,52) = 4.47, MS_e = 0.21$ ].

## DISCUSSION

We tested the hypothesis that LF words are better recognized than HF words because LF words contain more uncommon features than HF words. The results confirm the prediction made by the REM model of Shiffrin and Steyvers (1997; also Zechmeister, 1969, 1972): Words were better recognized when they consisted of primarily LF letters. The results also show that normative word frequency has an additional effect on recognition: LF words were recognized better than HF words even when their orthographic-feature frequencies were controlled. This suggests that orthographic-feature frequency is one, but not the only, factor underlying the WFE. It is possible (though there are good reasons to think other explanations play a role) that the entire WFE could be accounted for in terms of feature frequency if all the features that characterize a word could be manipulated (e.g., semantic, phonetic, etc.). However, this is probably not practical to do empirically. We note that the distinctiveness/feature-frequency account and the other accounts of the WFE are not necessarily mutually exclusive; it is entirely possible that more than one mechanism gives rise to the WFE.

If more than one mechanism gives rise to the WFE, the present findings may help us understand other accounts



**Figure 2.** The results of the experiment varying orthographic-feature frequency and normative word frequency. The results in terms of  $d'_a$  are shown in the upper panel, and the HRs (old) and FARs (new) are shown in the lower panel. Error bars are standard errors of the mean.

of the WFE better. For example, one popular theory of the WFE assumes that it is the result of more attentional resources being given to LF words than HF words when they are studied (e.g., Glanzer & Adams, 1990; Maddox & Estes, 1997). Within the context of that theory, the present findings suggest that more attention is given to words with distinctive features, and this produces the WFE.

Another theory of the WFE assumes that LF words are better recognized than HF words because HF words are encoded more variably than LF words (McClelland & Chappell, 1998). McClelland and Chappell assumed that this was because HF words tend to have more meanings than

LF words—and, therefore, subjects are more likely to attend to an LF feature than to an HF feature (on average). However, unless an argument is made for the role orthography plays in recognition memory, our findings are inconsistent with McClelland and Chappell's model.<sup>4</sup>

Dennis and Humphreys (2001) described another theory of the WFE, called BCDMEM, that is challenged by our findings. In BCDMEM, the WFE is predicted because HF words tend to appear in a larger number of contexts than do LF words (see Criss & Shiffrin, 2002); features that represent words play no role in producing the WFE. Here, we directly manipulated features that can only be attributed to the words, and, therefore, recognition should not have been affected according to the Dennis and Humphreys model. The hypothesis that contextual variability contributes to the WFE is not disconfirmed by our findings, since it may be that it operates in addition to distinctiveness. In fact, Shiffrin and Steyvers (1997) also proposed this but did not implement it in their REM models. Our findings are inconsistent with the hypothesis that contextual variability is the only factor underlying the WFE for recognition.

#### REFERENCES

- CRISS, A. H., & SHIFFRIN, R. M. (2002). *Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys*. Manuscript submitted for publication.
- DENNIS, S., & HUMPHREYS, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.
- GLANZER, M., & ADAMS, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8-20.
- GLANZER, M., & ADAMS, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 5-16.
- HUNT, R. R., & ELLIOTT, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, *109*, 49-74.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- JOORDENS, S., & HOCKLEY, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 1534-1555.
- 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 & Verbal Behavior*, *12*, 119-131.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MADDOX, W. T., & ESTES, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 539-559.
- MALMBERG, K. J., & MURNANE, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 616-630.
- MALMBERG, K. J., & SHIFFRIN, R. M. (2002). *The effect of study time on implicit and explicit memory: The "one-shot" hypothesis*. Manuscript submitted for publication.
- 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.
- SCHOOLER, L. J., SHIFFRIN, R. M., & RAAIJMAKERS, J. G. W. (2001). A Bayesian model for implicit effects in perceptual identification. *Psychological Review*, *108*, 257-272.
- SCHULMAN, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science*, *9*, 211-212.

- SHEPARD, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, **6**, 156-163.
- SHIFFRIN, R. M., & STEYVERS, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.
- SHIFFRIN, R. M., & STEYVERS, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73-95). Oxford: Oxford University Press.
- STEYVERS, M., MALMBERG, K. J., & SHIFFRIN, R. M. (2002). *The effect of contextual variability on recognition memory*. Manuscript submitted for publication.
- WIXTED, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 681-690.
- ZECHMEISTER, E. B. (1969). Orthographic distinctiveness. *Journal of Verbal Learning & Verbal Behavior*, **8**, 754-761.
- ZECHMEISTER, E. B. (1972). Orthographic distinctiveness as a variable in word recognition. *American Journal of Psychology*, **85**, 425-430.

### NOTES

1. In addition to features representing of word knowledge, contextual features are also assumed to be stored in lexical/semantic images (Schooler, Shiffrin, & Raaijmakers, 2001) and episodic images (Malmberg &

Shiffrin, 2002). In this highly simplified REM model, we do not take into account the role of context information at study or test (but see Shiffrin & Steyvers, 1997, and Steyvers, Malmberg, & Shiffrin, 2002). It is potentially important to consider the role context plays in producing the WFE (see Shiffrin & Steyvers, 1997, for a discussion). However, for the sake of simplicity, we omit its discussion in this short report.

2. A similar procedure could be used construct distinctive measures for phonological or higher order sublexical units.

3. We choose the procedure of selecting criteria separately for each subject for two different reasons. First, this procedure corrects for idiosyncratic use of the confidence scale (i.e., some subjects use one end of the scale more than do other subjects). Second, a subject-specific criterion leads to smaller standard errors in measures of sensitivity, hits, and false alarms than does a universal criterion.

An alternative procedure is to use one criterion for all subjects, such as the criterion between the first three and last three confidence ratings. With this alternative procedure, all statistical results remain the same: HRs were significantly higher for LF words than for HF words [ $F(1,52) = 10.36$ ;  $MS_e = 0.01$ ], and FARs were significantly lower for LF words than for HF words [ $F(1,52) = 23.44$ ,  $MS_e = 0.163$ ]. The HR for words with LF letters was slightly greater than the HR for words with HF letters features [ $F(1,52) = 1.21$ ,  $MS_e = .01$ ], and the FARs were significantly lower for words with LF letters than for words with HF letters [ $F(1, 52) = 50.76$ ,  $MS_e = 0.43$ ].

4. We thank Mark Chappell for pointing this out.

### APPENDIX A Words in the Four Conditions

| Low Letter Frequency and Low Word Frequency  |         |         |         |         |         |         |         |
|--|---------|---------|---------|---------|---------|---------|---------|
| ABLAZE                                       | CHIMP   | ERGO    | JAGGED  | LIEU    | OPOSSUM | QUICKEN | TYPHOID |
| ACRYLIC                                      | CHOMP   | EXCERPT | JOGGING | LOCKS   | OUTBACK | QUIP    | UPTIGHT |
| AJAR   | CHUBBY  | EXHALE  | JOWL    | LYRICS  | OUTGROW | QUIRK   | UTOPIA  |
| ALFALFA                                      | CONVEX  | EXHAUST | JUNO    | MAYFLY  | OZONE   | REVAMP  | VERB    |
| APEX   | DYNAMIC | FLUX    | KILO    | MIDRIFF | PREFIX  | SKIMP   | VIVA    |
| AVOCADO                                      | ELYSIUM | GAWKY   | KIOSK   | NOVA    | PSYCHE  | SQUID   | VORTEX  |
| AVOW   | ENCAMP  | GUSTO   | KNACK   | NUMBLY  | PUFFY   | STANZA  | WHACK   |
| AZALEA                                       | EPIC    | HUMP    | KNOBBLY | ODYSSEY | QUAKE   | SWAB    | YANK    |
| BOXING                                       | EPOCH   | IMPEL   | KNOWING | OOZE    | QUIBBLE | TWITCH  | YOLK    |
| High Letter Frequency and Low Word Frequency |         |         |         |         |         |         |         |
| ALERT  | BROILER | CURLY   | FAINT   | PARROT  | PETITE  | SEARING | SOLID   |
| BANE   | BRUTE   | CURRENT | FERRET  | PASTE   | PLIANT  | SEDATE  | SOOT    |
| BARTER                                       | CALLER  | DALE    | FLIER   | PATE    | PORE    | SENSORY | SPORE   |
| BASTE  | CENSURE | DEAREST | GALORE  | PATRIOT | RELIANT | SHEAR   | STERIOD |
| BEET   | COERCE  | DECREE  | LEARNER | PEAT    | RILE    | SHINE   | STRUT   |
| BILE   | COOLER  | DELETE  | MANE    | PELLET  | SAIL    | SILT    | SUNRISE |
| BOILER                                       | CORNET  | DILATE  | MARINER | PENAL   | SAUCY   | SINNER  | TANNERY |
| BRAID  | CORONER | DINER   | MIRE    | PENANCE | SAUNTER | SMEAR   | TENSE   |
| BRAY   | COTE    | DIRE    | PALETTE | PERT    | SCARLET | SNOOTY  | TINE    |
| Low Letter Frequency and High Word Frequency |         |         |         |         |         |         |         |
| AMAZING                                      | DOZEN   | EXPLODE | KICK    | MAJOR   | OTTO    | TAXI    | UNIQUE  |
| ATOMIC                                       | EGYPT   | EYEBROW | KINGDOM | MIXED   | OXYGEN  | THIGH   | UNKNOWN |
| AWFULLY                                      | ELBOW   | GHETTO  | KNIGHT  | MYTH    | PHOTO   | THOU    | UPWARDS |
| AWKWARD                                      | EVOLVE  | GOLF    | LAMB    | NATO    | PHYSICS | THUMB   | VACUUM  |
| BUREAU                                       | EXAM    | GULF    | LIMB    | NETWORK | PUZZLED | TOBACCO | WAYS    |
| CLIFF  | EXCEED  | HAZARD  | LIQUID  | ODDS    | RHYTHM  | TOMB    | WHIP    |
| CLIMB  | EXCLAIM | INDEX   | LOBBY   | OFFEND  | RUBBER  | UNDERGO | WHISKY  |
| COMPLEX                                      | EXERT   | INJURY  | LOGIC   | OMEGA   | SYMBOL  | UNHAPPY | WIDOW   |
| DIFFER                                       | EXIT    | JACKAL  | LUXURY  | OPERA   | SYMPTOM | UNIFORM | ZERO    |

**APPENDIX A (Continued)**

High Letter Frequency and High Word Frequency

|         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|
| AIRLINE | BLEED   | CURE    | GREET   | PENALTY | POLE    | SEAL    | STRAIN  |
| BAIT    | CANAL   | CURRENT | MALE    | PILE    | PRAY    | SECURE  | TALE    |
| BALLET  | CATTLE  | DAISY   | MINER   | PILOT   | PRESENT | SENATOR | TENURE  |
| BARREL  | CELLAR  | DEALER  | MINERAL | PINE    | RALLY   | SHEER   | TERRACE |
| BARRIER | CLAY    | DENSE   | MIRACLE | PLAIN   | RELATE  | SHORE   | TERROR  |
| BEAR    | CLIENT  | FARE    | PAINTER | PLANET  | RELEASE | SPINE   | TOILET  |
| BEAST   | CORE    | FLEET   | PANEL   | PLANNER | RETIRE  | STARTLE | TRACE   |
| BETRAY  | CORRECT | FREE    | PARADE  | PLEAD   | SAME    | STATUE  | TRAY    |
| BITE    | CRUELTY | GALLERY | PEASANT | POET    | SCENT   | STORAGE | TREATY  |

**APPENDIX B  
Means and Standard Deviations of the Word Frequencies  
and Letter Frequencies**

| Word Frequency        | Letter Frequency |           |          |           |
|-----------------------|------------------|-----------|----------|-----------|
|                       | Low              |           | High     |           |
|                       | <i>M</i>         | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Word Frequency        |                  |           |          |           |
| Low                   | 4.1              | 1.2       | 4.6      | 1.4       |
| High                  | 23.6             | 6.6       | 25.3     | 7.2       |
| Mean Letter Frequency |                  |           |          |           |
| Low                   | .052             | .012      | .095     | .012      |
| High                  | .054             | .011      | .094     | .013      |

(Manuscript received October 3, 2001;  
revision accepted for publication February 12, 2002.)