Journal of Memory and Language 59 (2008) 331–345

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency

Amy H. Criss [a,*], Kenneth J. Malmberg [b,*]

[a] Department of Psychology, Syracuse University, 477 Huntington Hall, Syracuse, NY 13210, USA
[b] Department of Psychology, University of South Florida, 4202 East Fowler Ave, PCD 4118G, Tampa, FL 33620, USA

ARTICLE INFO

Article history:
Received 1 August 2007
revision received 22 May 2008
Available online 14 July 2008

Keywords:
Episodic memory
Word frequency effect
Memory models
Object memory
Orthography

ABSTRACT

One of the most studied and least well understood phenomena in episodic memory is the word frequency effect (WFE). The WFE is expressed as a mirror pattern where uncommon low frequency words (LF) are better recognized than common high frequency words (HF) by way of a higher HR and lower FAR. One explanation for the HR difference is the early-phase elevated-attention hypothesis which proposes two stages of encoding. In the first, called the early-phase, words are identified based on orthographic and/or phonological characteristics. LF words are composed of atypical features making their identification more difficult than HF words. This relative difficulty during the early-phase results in the LF HR advantage. The first two experiments test the proposal that LF words are better recognized due to their distinct lexical features. The second stage of encoding, called the late phase, consists of controlled processing where the semantic features of the item are paramount. According to the early-phase elevated-attention hypothesis, semantic features of HF and LF words do not differ in diagnosticity and do not contribute to the word frequency effect. We find evidence for this assumption in the final experiment by comparing memory for words and objects.

The occurrences of uncommon words are better recognized than the occurrences of common words (Glanzer & Adams, 1985; Schulman, 1967; Shepard, 1967): The hit rate (HR) is greater and the false alarm rate (FAR) is lower for uncommon words. Despite over 40 years of research, there is no consensus on how to explain the word frequency effect (WFE). The WFE has been attributed to a variety of different factors or mechanisms (Criss & Shiffrin, 2004a; Dennis & Humphreys, 2001; Glanzer & Adams, 1990; Malmberg & Murnane, 2002; Malmberg, Steyvers, Stephens, & Shiffrin, 2002; McClelland & Chappell, 1998; Murdock, 2003; Reder et al., 2000; Shiffrin & Steyvers, 1997; Steyvers & Malmberg, 2003). The focus of this article is the elevated-attention hypothesis, which proposes that a differential amount of attention is devoted to high frequency (HF) vs. low frequency (LF) words, with more attention given to the latter, giving rise to the word frequency effect (Brown, 1976; Glanzer & Adams, 1990; Lockhart, Craik, & Jacoby, 1976; Malmberg & Nelson, 2003; Shepard, 1967). We are specifically concerned with the type of information that gives rise to WFE.

The Attention Likelihood Theory (ALT; Glanzer & Adams, 1990) is one well-known elevated-attention model. ALT assumes that LF words attract more attention during encoding than HF words, and thus a greater number of features are "marked" in a LF memory trace than in a HF memory trace. This produces the LF HR advantage. The LF FAR advantage is produced by a meta-cognitive judgment that discounts a certain number of marked features for LF words on the expectation that the trace of a recently studied LF word should contain a relatively large number of marked features if it were studied. Thus, the system

* Corresponding authors.
E-mail addresses: amy.criss@gmail.com, acriss@syr.edu (A.H. Criss), malmberg@cas.usf.edu (K.J. Malmberg).

0749-596X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2008.05.002

requires more evidence to respond positively to LF than to HF words.[1]

One of the critical predictions of ALT is that factors that influence the mirror patterned WFE should have equal impact on HRs and FARs. Under no condition should a LF HR advantage be observed in the absence of a LF FAR advantage and vice versa. Other versions of the elevated-attention hypothesis predict that only the magnitude of LF HR advantage should be positively related to the amount of attentional resources available during study. To test this prediction, Malmberg and Nelson (2003) varied study duration on the assumption that increases in study time would result in increases in the amount of attentional resources available during encoding. The result was that increasing study time increased the LF HR advantage when study durations were relatively short (i.e., .25–1.0 s) but not when study times were relatively long (i.e., greater than 1.0 s). At the shortest study time, the LF HR advantage was not observed, but an intact LF FAR advantage was always observed. This complex relationship between study time and the LF HR advantage was difficult for ALT and other versions the elevated-attention hypothesis to explain.

Malmberg and Nelson (2003) also explored how a different form of attentional load affected the WFE by having subjects study pairs of words and testing memory with single items. In accordance with the elevated-attention assumption they assumed that LF words attract more attention than HF words on average. Hence, words studied with LF words should receive less attentional resources during study than words studied with HF words. If the LF HR advantage results from extra attention devoted to LF words, the effect should be disrupted when attentional resources are taxed. Their results demonstrated poorer performance for words studied with LF words than words studied with HF words only when study time was limited to 800 ms per pair. When 4000 ms of study time per pair was available, pair-type had no effect. Critically, no LF HR advantage was observed in the 800 ms study condition where attentional resources were taxed, but the LF HR advantage was observed in the 4000 ms study time condition where attentional resources were not fully taxed. Despite the variable pattern of LF HR advantages, an LF FAR advantage was consistently observed. These results disconfirmed the ALT account of the WFE for two reasons: The LF FAR advantage was observed in the absence of the LF HR advantage and the LF HR advantage was only observed for study times greater than 800 ms.

The early-phase elevated-attention hypothesis

To account for these findings, Malmberg and Nelson (2003) proposed the *early-phase elevated-attention hypothesis* (EPEA). ALT makes no distinction between features representing lexical aspects vs. semantic aspects of words. In contrast, the early-phase elevated-attention hypothesis assumes two successive phases of encoding a word into episodic memory. During the early-phase, words are identified based on phonological and/or orthographic characteristics. As a result, the meaning of the word is retrieved and held in short term/working memory. During the late phase, participants actively control the processing of higher level semantic features of the words held in short term memory. Encoding strategies such as rehearsal, sentence formation, imagery, and others are assumed to occur during this later controlled phase of encoding (Atkinson & Shiffrin, 1968; Nelson & Narens, 1990). Accordingly, the EPEA hypothesis assumes that the LF HR advantage is due to increased attentional resources allocated to LF words relative to HF words during the early-phase in which the word is identified, and the late phase does not contribute to the LF HR advantage.

A critical assumption is that uncommon words are more difficult to identify than common words. Evidence that LF words are more difficult to process or attract more attention than HF words comes from several types of experiments. Some show that identification latencies are longer for LF than HF words (e.g., Besner & McCann, 1987; Borowsky & Besner, 1993; Forster & Chambers, 1973; Scarborough, Cortese, & Scarborough, 1977), others show that naming a picture paired with an unrelated word is slower when that word is LF rather than HF (Miozzo & Caramazza, 2003), and still other experiments show that performance is less accurate or slower on a secondary task while studying LF words compared to when HF words are studied (Naveh-Benjamin, Craik, Guez, & Dori, 1998; Naveh-Benjamin & Guez, 2000).

The relative difficulty in identifying LF words and the subsequent benefit in recognition memory may result from the uncommon structural aspects of LF words. Evidence that the structure of LF and HF words differs comes from Landauer and Streeter (1973). They found that HF words have more neighbors (words that differ by a single letter) and those neighbors tend to be of higher frequency than the neighbors of LF words. They also found that LF words are composed of longer phonemes than HF words (holding word length constant) and that the distribution of letters is different for HF and LF words (e.g., the letters T, L, and E are relatively more common in HF words and the letters S, P, and A are relatively more common in LF words). The uncommon structural aspects of LF words may aid recognition memory. Words with uncommon letters are recognized better than words with common letters, even when normative word frequency is held constant (Malmberg et al., 2002) and words rated by participants as orthographically distinct are better recognized than words rated less distinct (Zechmeister, 1972).

According to the EPEA hypothesis, the relative difficulty in processing LF words requires additional resources during encoding which results in a more complete, accurate, and/or more diagnostic trace stored in episodic memory. This in turn increases the ability to later recognize a word that was more difficult to initially process. Bjork (1994) proposed that factors that make learning more difficult lead to better memory performance. For instance, generating items vs. reading items requires differences in effort during study and this often enhances memory (Slamecka

[1] Several of ALT's assumptions have been refuted (e.g., Balakrishnan & Ratcliff, 1996; Criss & McClelland, 2006; Hintzman, 1994; Malmberg & Murnane, 2002), but other variants of the elevated-attention hypothesis remain viable.

& Graf, 1978; Westerman & Greene, 1997, but also see Serra & Nairne, 1993). Additionally, post-masking a brief presentation of word actually improves recognition memory (Hirshman & Mulligan, 1991; Hirshman, Trembath, & Mulligan, 1994; Mulligan, 2000). Thus, there is convergent support for the assumption that recognition memory is improved by making words more difficult to identify.

In sum, there is evidence that an early-phase of encoding, during which word identification occurs, is responsible for a substantial portion of the LF HR advantage in episodic memory. The assumption that the later phase of encoding does not typically contribute to the LF HR advantage is a bit more speculative at this time. Malmberg and Nelson (2003) proposed that the semantic and associative features stored during this phase are similarly distinct for LF and HF words. For instance, *trout* and *halibut* vary in the rate with which they occur in everyday life, but they are likely to invoke similar thoughts having to do with water, fishing, dinner, etc. However, if more is typically known about HF words than LF words, it is even possible that the late phase could favor HF words due to the ability to store additional (but not more diagnostic) information about them. If so, then operations that enhance semantic encoding might disrupt and possibly reduce the LF HR advantage.

To test this assumption, Criss and Shiffrin (2004a) had participants perform a wide variety of different orienting tasks during study (e.g., Craik & Tulving, 1975). All tasks, with the exception of one, eliminated or drastically reduced the LF HR advantage (for similar results see Glanc & Greene, 2007; Guttentag & Carrol, 1997; Hirshman & Ardnt, 1997; Mandler, Goodman, & Wilkes-Gibbs, 1982). The exception was a task that required a judgment about whether or not the word contained any unusual letters. In this condition and when participants are left to their own devices with no assigned task, the LF HR advantage was observed. Criss and Shiffrin (2004a) also manipulated study time. Their findings support the assumption of the EPEA hypothesis that the LF HR advantage should emerge in first second or so of study time when it is observed and it should not increase in magnitude with greater amounts of study time. Together, these findings suggest that participants encode task relevant features during the late phase of encoding. When engaged in semantic tasks (i.e., concreteness task) they store additional semantic features. When engaged in tasks focused on orthography (i.e., unusual letters task) they store additional orthographic features reintroducing a higher HR for LF words.[2]

This paper consists of experiments addressing both the early- and late-phases of processing outlined above. The first section tests the idea that the LF HR advantage results from the unusual features present in LF words (i.e., letters, bigrams, trigrams, etc.) making them more difficult to process and identify during the early-phase of encoding. To the extent that the LF HR advantage results from differences in the amount of attention required to identify words, the effect can be eliminated by holding orthography/phonology constant. Then we consider the assumption that semantic and associative features that are the focus of the later phase of encoding do not differ in diagnosticity as a function of normative word frequency. Hence, when encoding of the orthographic/phonological information associated with lexical access is rendered unimportant or unnecessary, the LF HR advantage should also be eliminated.

Experiment 1

In this experiment, letter frequency, normative word frequency, study time, and the orienting task are manipulated within a single study list. We operationally define orthography/phonology as letter frequency (see Malmberg et al., 2002, for a complete description of the measure and a table of relative frequencies of letter by position in the word).[3] The origin of this measure is the Retrieving Effectively from Memory (REM) model of episodic memory (Shiffrin & Steyvers, 1997). In REM, the amount of positive evidence provided by a match is determined by the diagnosticity of the particular matching feature. An uncommon feature provides more evidence that the test item was studied because matching an uncommon feature is unlikely to occur by chance. Malmberg et al. tested this principle and found that diagnosticity was positively related to recognition memory performance. For orienting task, we include the unusual letters task and no task because both induce a LF HR advantage when letter frequency is not controlled. We also include the concreteness task, one of the many tasks that eliminate the LF HR advantage (Criss & Shiffrin, 2004a).

Predictions of the early-phase elevated-attention hypothesis

To the extent that the benefit for LF targets in prior studies was the result of the extra attention allocated to them based on their orthographic distinctiveness, then there should be an advantage for targets composed of distinct letters. Based on prior results and assumptions of the EPEA hypothesis, we also predict that the benefit of distinct letters will emerge within one second and the magnitude of that benefit will not grow with additional encoding time (e.g., Criss & Shiffrin, 2004a; Malmberg & Nelson, 2003). The strictest interpretation of the EPEA hypothesis predicts no difference in the attentional resources devoted to encoding LF and HF words when letter frequency is held constant.

Moreover, the predictions hold for all orienting tasks. According to the EPEA hypothesis, the effect of letter frequency is obligatory and results from word identification. Thus, the effect of letter frequency should not be sensitive to the demands of the orienting task during the late phase of controlled processing. When letter frequency was not

[2] We are not proposing that orthography/phonology is to equivalent word frequency nor are we claiming that manipulations of orthography/phonology must demonstrate similar effects as WF.

[3] The early-phase elevated-attention hypothesis is not committed to a single measure of orthography such as the letter frequency measure we employ. Rather, the hypothesis is more general and claims that difficult orthographic/phonological properties require extra attention to identify resulting in an episodic memory trace that is more accurate and/or diagnostic enhancing recognition memory. We assume that one factor contributing to word identification during the early-phase of encoding is letter frequency. To the best of our knowledge, this operational definition of orthography has not been explored in the word recognition literature.

controlled, Criss and Shiffrin (2004a) showed that LF HRs are greater than HF HRs when encoded with the unusual letters task but LF and HF HRs are not different for other orienting tasks. They attributed the equal HRs to task demands requiring processing of semantic features during the late phase. They attributed the reintroduction of the LF HR advantage with the unusual letters task to the encoding of orthography during the late phase of controlled processing. In this experiment where letter frequency is held constant for LF and HF words, directing attention to orthography will not differentially benefit LF words because they are now equivalent on this factor. Thus, the prediction is that neither word frequency nor letter frequency will interact with the orienting task.

Method

Participants

Ninety people from the Indiana University or Carnegie Mellon University communities participated in the experiment in exchange for partial course credit or $7.00 per hour.

Materials

The word pool consisted of 4 sets of 72 words taken from Malmberg et al. (2002). Normative word frequency and normative letter frequency (which is a function of both the letter itself and its position in the word; see the original source for additional details including the actual stimuli) were orthogonally crossed to form the word sets, see Table 1. The HF words averaged 24.45 times per million (range 15–39) and the LF words averaged 4.35 times per million (range 3–7) in the Celex database (Baayen, Piepenbrock, & Gulikers, 1995). This range of word frequencies is known to produce the typical mirror patterned word frequency effect (e.g., Estes & Maddox, 1995). The words with distinct letters had a relative mean letter frequency of .053 and the words with typical letters had a mean letter frequency of .0945, and this range of letter frequencies is known to produce a mirror effect (Malmberg et al., 2002). Words within each set were assigned to be a foil or assigned to one of the target conditions randomly for each participant.

Procedure and design

Word frequency (HF and LF), letter frequency (distinct and typical), study time (0.5, 1, and 3 s), and orienting task (concreteness, unusual letters, and no task) were manipulated within participant, within list. The concreteness task asks, “Does this word represent something you can see, hear, taste, smell, or feel?” The unusual letters task asks, “Does this word contain any unusual letters?” The no task condition does not require any response from the participant. In addition to the question appearing on the screen, the background color of the screen and the color of the word varied with orienting task as an additional cue to perform the appropriate task. Each participant was informed that an unspecified memory test would follow the study list.

The study list consisted of 144 words, equally divided among the 36 conditions. The order of trials was randomly assigned for each participant. On each study trial, a blank screen appeared for 150 ms followed by a screen containing the study word and the question to be answered (trials in the no task condition did not include a question). The response could be entered at any time during the trial, however, the screen did not change until the appropriate amount of study time passed and then the next trial began. At the end of the study list, participants engaged in 30 s of math problems followed by 288 individual test items. The test was a yes–no recognition memory test and consisted of all 144 targets (an equal number from each condition) along with an equal number of foils, randomly intermixed. The foils consisted of an equal number of words from the 4 conditions crossing word frequency and letter frequency. An alpha level of .05 was used for all statistical tests.

Table 1
Mean values of normative word frequency and letter frequency

Word frequency condition	Letter frequency condition	
	Distinct	Typical
	Letter frequency	
Low	.052 (.012)	.095 (.012)
High	.054 (.011)	.094 (.013)
	Word frequency	
Low	4.1 (1.2)	4.6 (1.4)
High	23.6 (6.6)	25.3 (7.2)

Note. Standard deviations are in parentheses. Adapted from Appendix B of Malmberg et al. (2002).

Results and discussion

The HRs are shown in Fig. 1. A 2 × 3 × 3 × 2 repeated measures analysis of variance (ANOVA) was conducted on the HRs (letter frequency, study time, orienting task, and word frequency). Neither word frequency nor orthographic distinctiveness interacted with orienting task, as predicted. There was a reliable interaction between study time and orienting task, $F(4,356) = 3.12$, $MSE = .047$, $p = .015$, due to the slightly more shallow increase in HR as a function of study time for the no task condition compared to the other encoding tasks. No other interactions approached significance.

HRs for words with distinct letters were greater than HRs for words with typical letters, $F(1,89) = 6.16$, $MSE = .079$, $p = .015$. HRs were positively related to study time, $F(2,178) = 67.53$, $MSE = .066$, $p < .001$. There was a main effect of orienting task, $F(1,178) = 77.23$, $MSE = .078$, $p < .001$. The effect of orienting task was further analyzed with Bonferroni adjusted post-hoc tests showing that HRs are ordered as follows concreteness > unusual letters > no task. Importantly, there was not a reliable effect of word frequency, $F < 1$.

In accord with the predictions, targets composed of distinct letters were better remembered than words composed of typical letters for all orienting tasks, and this effect was present within the first second of encoding and did not change in magnitude thereafter (i.e., there was no encoding time by letter frequency interaction). This novel finding provides evidence that letter frequency contributes to word identification and is therefore not sensi-

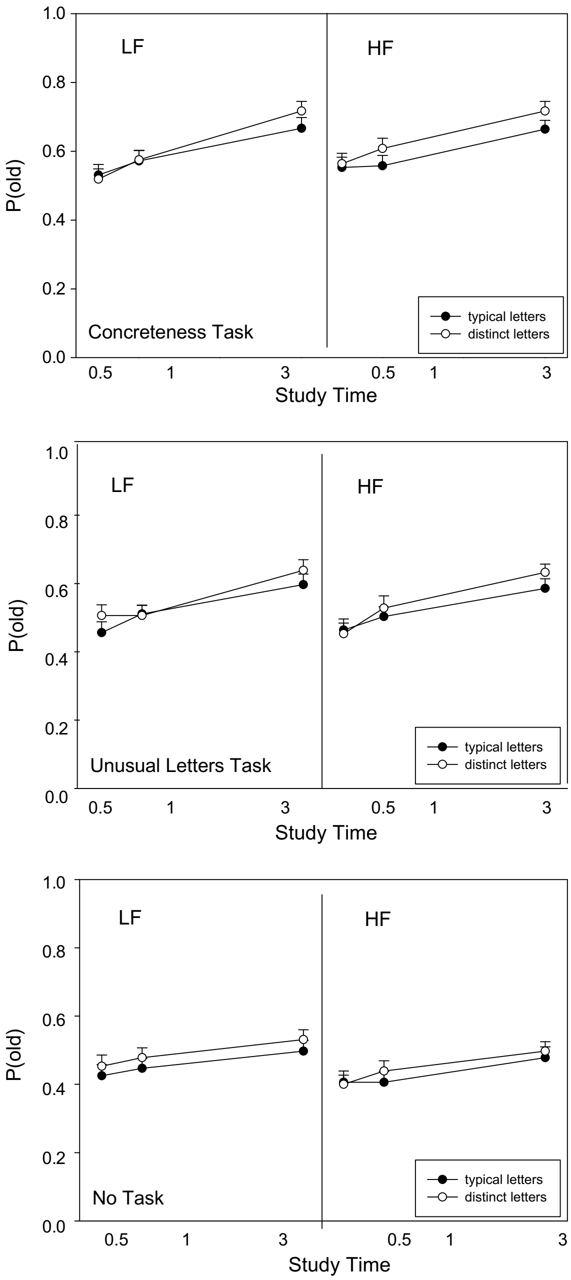

Fig. 1. Hit rates (HR) for words in Experiment 1 plotted separately for each encoding task. The top panel shows performance for words encoded with the concreteness task, the middle panel shows performance for the unusual letters task, and the bottom panel shows performance for words encoded with out any defined task. Low frequency (LF) HRs are plotted on the left and high frequency (HF) HRs are plotted on the right. Words composed of distinct letters are show as white circles and words composed of typical letters are shown as black circles.

tive to task demands and other processes that take place during the late phase of encoding.

There was no difference between LF and HF HRs and the orienting task by WF interaction approached but did not achieve significance, $F(2,178) = 2.41$, $MSE = .065$, $p = .093$. Nevertheless, we examined simple main effects of WF separately for each orienting task. Replicating prior findings and consistent with our predictions, we did not find a reliable difference in HRs for the concreteness task, $F(1,89) = 1.12$. Consistent with the EPEA hypothesis, we found no difference between LF and HF HRs for words encoded with the unusual letters task, $F < 1$, probably because we controlled for that dimension. Inspection of Fig. 1 suggests a greater HR for LF than HF words in the no task condition, consistent with the findings of Malmberg et al. (2002). However, this effect is not reliable, $F(1,89) = 3.51$, $MSE = .045$, $p = .064$.

The FARs are listed in Table 2. A 2 × 2 repeated measures ANOVA (word frequency and letter frequency) indicates that FARs were greater for HF words than LF words, $F(1,89) = 4.72$, $MSE = .009$, $p = .032$, and FARs were greater for words with typical letters than for words with distinct letters, $F(1,89) = 16.09$, $MSE = .009$, $p < .001$. The interaction was not significant, $F < 1.0$.

Experiment 2

The goal of this experiment is to replicate the critical conditions of Experiment 1 and to obtain additional data for the no task condition. In this experiment, normative WF, letter frequency, and orienting task (letter task and no task) were manipulated.

Method

Participants

Fifty undergraduates at Syracuse University participated to fulfill a course requirement. Two failed to follow instructions leaving 48 participants who contributed data to the analyses reported below.

Materials

The materials were identical to Experiment 1.

Procedure and design

Word frequency (HF and LF), letter frequency (distinct and typical), and orienting task (unusual letters and no task) were manipulated within participant, within list. The study list consisted of 144 words equally divided among the 8 conditions. Each study trial consisted of a blank screen for approximately 150 ms followed by the study word which appeared for 1.5 s during which time the participant completed the unusual letters task or engaged in no task. The test list consisted of all targets and an equal number of foils randomly intermixed anew for

Table 2
Mean and standard errors of the mean for false alarms to low and high frequency words as a function of letter frequency

	Word frequency	Letter frequency	
		Distinct	Typical
Experiment 1	Low	.248 (.021)	.292 (.020)
	High	.273 (.020)	.311 (.020)
Experiment 2	Low	.192 (.018)	.235 (.022)
	High	.208 (.021)	.273 (.027)

each participant. All remaining details were identical to Experiment 1.

Results and discussion

HRs are plotted in Fig. 2. A 2 × 2 × 2 repeated measures ANOVA on HRs (letter frequency, orienting task and word frequency) revealed a reliable three-way interaction, $F(1,47) = 7.21$, $MSE = .011$, $p = .01$ thus we consider the two tasks separately. The results of the unusual letters task condition replicated the results of Experiment 1: HRs were significantly greater for words composed of distinct letters than for words composed of typical letters, $F(1,47) = 15.76$, $MSE = .015$, $p < .001$, and there was a null effect of word frequency, $F(1,47) = 2.41$. The interaction was also unreliable, $F(1,47) = 2.12$. A 2 × 2 repeated measures ANOVA on HRs in the no task condition revealed a main effect of letter frequency that approached significance, $F(1,47) = 3.40$, $MSE = .015$, $p = .071$, and no reliable effect of word frequency, $F(1,47) = 1.48$. There was an interaction between word frequency and letter frequency, $F(1,47) = 4.80$, $MSE = .012$, such that HRs were greater for LF words composed of distinct letters than typical letters, $t(47) = 2.67$, $p = .01$, but no difference for HF words, $t(47) = .05$.

The FARs are shown in Table 2. A 2 × 2 repeated measures analysis of variance (ANOVA) (word frequency and letter frequency) replicated Experiment 1 with a greater FAR for HF than LF words, $F(1,47) = 16.96$, $MSE = .008$, $p < .001$ and a greater FAR for words with typical than distinct letters, $F(1,47) = 6.75$, $MSE = .005$, $p = .012$. The interaction was not significant, $F(1,47) = 2.18$.

Discussion of Experiments 1 and 2

The results of Experiments 1 and 2 indicate that orthographic aspects of words play an important role in recognition memory (Criss & Shiffrin, 2004a; Estes & Maddox, 2002; Glanc & Greene, 2007; Hunt & Elliott, 1980; Malmberg et al., 2002; Shiffrin & Steyvers, 1997; Zechmeister, 1972). The persistent distinctive letter HR effect across all orienting tasks and its emergence within the first second of study confirm predictions of the EPEA hypothesis. Words are identified before a later controlled phase of encoding occurs. Overcoming the difficulty associated with identifying words with uncommon orthography demands attention, and thus enhanced encoding. Since the effect of letter frequency occurs during the early-phase of word identification, it is not subject to control by the participant and occurs independently of the attention directed to various subsets of features during the later stages of processing.

When certain orienting tasks are used during study the LF HR advantage is eliminated. Here, this occurred with the use of a concreteness task, replicating the results of prior experiments (Criss & Shiffrin, 2004a; Hirshman & Arndt, 1997). Perhaps more importantly, the LF HR advantage was eliminated when we controlled for orthographic distinctiveness even when participants were engaged in a task focused on unusual letters. In prior experiments, letter frequency was not controlled and the LF HR advantage was observed for words encoded with this task (Criss & Shiffrin, 2004a). There, focused attention on orthography during the late stage of encoding could have preserved the LF HR advantage. It is interesting to note that attending to orthography benefits encoding; HRs are greater for the unusual letters task than no task. One possibility is that the storage of new orthographic features prior to the controlled phase of encoding increased the vigilance of participants during study.

When no orienting task is specified and letter frequency is controlled, the results are more variable over all. In one published experiment, LF targets are better remembered than HF targets (Malmberg et al., 2002), marginal trends were observed in Experiment 1, but there was no LF HR advantage in Experiment 2. This makes sense, since we do not know (or at least cannot be reasonably certain) what participants are attending to when left to their own devices and they may have idiosyncratic encoding strategies or even lapses in attention. What is striking, however, is that these studies all demonstrate an effect of letter distinctiveness. The reliability of letter frequency effects and transient effects of WF in these studies suggests that WFEs might be influenced by subtle differences in samples or procedures.

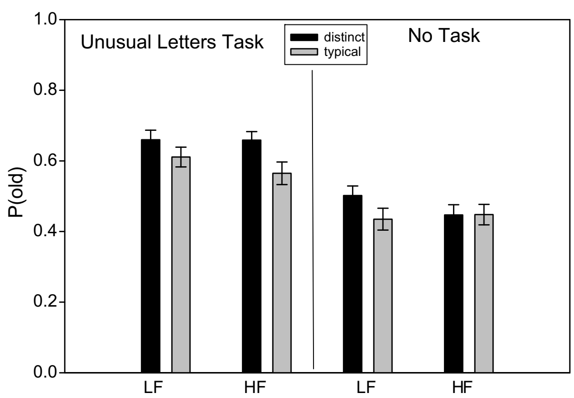

Fig. 2. HRs for words in Experiment 2. The left panel shows the hit rates (HR) for targets encoded during the unusual letters task. The right panel shows the HRs for targets encoded without any orienting task. Letter frequency (typical or distinct) and word frequency (LF, low frequency; HF, high frequency) were orthogonally manipulated.

Experiment 3

In the prior experiments, we varied orthographic distinctiveness and orienting task in ways that would affect the LF HR advantage as predicted by the EPEA hypothesis. In the next two experiments, we explored what would happen if we eliminated orthographic information as a source of mnemonic evidence. Malmberg and Nelson (2003) assumed that the semantic characteristics of words do not contribute to the LF HR advantage. To test this, we compare recognition performance for objects and their dominant labels which vary in normative frequency (Snodgrass & Vanderwart, 1980).

A number of studies suggest that words and objects are encoded in different ways (see Nelson, 1979, for a review; Martin & Wiggs, 1997). This might be due to how their physical characteristics are processed and represented in memory. Consider that certain features of words are invariant from encounter to encounter (e.g., orthography and phonology). Moreover, these relatively low level features vary in how often they are encountered, and thus some features or combinations of features may attract more attention in order to identify the words that they comprise (Malmberg et al., 2002). On the other hand, the features of any given object can be combined in an infinite number of ways, and thus the visual system must be able to identify objects from a potentially infinite number of view points (Biederman, 1987; Hayward & Tarr, 1997; Hummel & Biederman, 1992; Poggio & Edelman, 1990). How object recognition occurs is a controversial topic, but the ability to recognize objects from a variety of viewpoints is not. The high variability in how objects are viewed is in stark contrast to the stability with which words are viewed. Because the features of the same object may be viewed in a variety of different ways, object features probably do not vary in frequency of occurrence in the same way that orthographic features vary in frequency of occurrence.

The sensory/semantic model was developed by D.L. Nelson and colleagues (e.g., Nelson, 1979; Nelson & Brooks, 1973; Nelson, Reed, & McEvoy, 1977) to account for several recall findings that were difficult to explain by the then dominant dual code theory (Paivio, 1971). To the best of our knowledge the sensory/semantic model has never been applied to recognition memory. Here we do so by combining it with the EPEA hypothesis. The sensory/semantic model is depicted in Fig. 3, and it makes four assumptions:

1. There are different types of features that can be used to encode an object or a word: visual, orthographic, phonological, and meaning. The meaning features are assumed to be functionally the same regardless of whether they are accessed as the result of the presentation of an object or a word.
2. There are constraints on the order in which these different features are accessed. Assuming a visual stimulus presentation, both visual and orthographic features can directly access meaning features. Alternatively, and perhaps in parallel, the meaning features may be accessed via phonological features when a word is presented, but not when an object is presented. In order to access the phonological or orthographic features associated with an object, the object's meaning features must first be accessed.
3. Attention can be directed to a subset of the different types of features, which will affect the types of features that are encoded (cf., Criss & Shiffrin, 2005; Underwood, 1969).
4. More than one type of feature can be encoded; the more types of features encoded and used during retrieval, the more distinctive the resulting trace will be. Distinctiveness is assumed to be positively related to memory performance.[4]

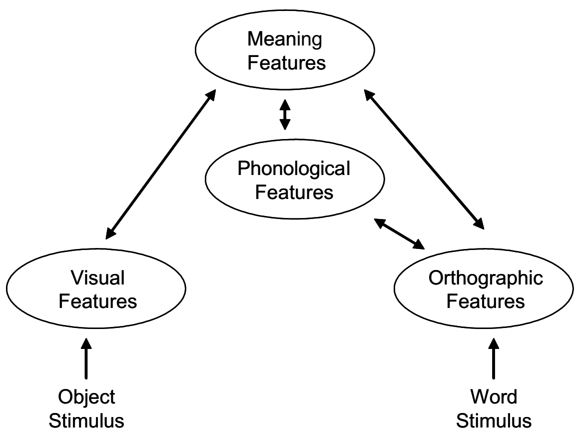

Fig. 3. The sensory/semantic model of D.L. Nelson and colleagues.

When an object is studied, visual and meaning features are stored. The model predicts no WFE when objects are studied because we assume that neither visual object features nor semantic features differ in diagnosticity. One difference between encoding a word and an object is the availability of orthographic features during initial encoding. LF words are more difficult to identify due to their unusual orthographic properties (i.e., as reviewed earlier in this paper). Perhaps, in addition, the direct path from orthographic features to meaning is less likely to be pursued for LF words and the phonologically mediated path might be more likely to be pursued. If so, phonological features would be more likely to be stored for LF than HF words, which would lead to more types of features stored and a more diagnostic LF episodic trace according to the sensory/semantic model. Lastly, the additional resources required to identify a LF word (i.e., access its meaning) via either pathway might increase the probability of storing orthographic features, which we assume gives rise to the LF HR advantage.

Merging the sensory/semantic model with assumptions from the EPEA hypothesis predicts the standard LF HR advantage only when words are studied and words are tested because only in this case are the orthographic features present in both the memory trace and test probe. We expect no difference between the LF and HF HR when objects are studied and tested because orthographic features are unlikely to play a role. This prediction does not require the strong assumption outlined above that the features of objects do not vary in normative frequency. One need merely assume that the normative frequency of individual features of the object is not correlated with letter frequency of the object's label.[5] Likewise, in the mixed conditions (words are studied and objects are tested or objects are studied and words are tested) we expect no LF HR advantage because these decisions primarily rely on seman-

[4] The authors usually mean that features themselves vary in distinctiveness. But Nelson et al. mean that encoding many different types of features makes a trace less confusable with other traces. For instance, ball and ball can be distinguished based only on the visual features associated with sports versus dances.

[5] We thank an anonymous reviewer for pointing this out.

tic features which we assume do not differ in diagnosticity as a function of frequency.

In the next experiment, participants studied either objects or their single word labels that varied in normative frequency and are then tested with words or objects. An auxiliary assumption to the predictions just outlined is that the frequency with which words occur in the natural language is correlated with the frequency with which the whole object occurs in the environment. To address the validity of this assumption, we first conducted an experiment in which a list of objects was studied, and memory for these objects was tested via free recall. The critical question is whether variations in normative frequency of the object's to-be-recalled label would produce the standard findings of better recall of HF words.[6]

Experiment 3a

Methods

Participants

Thirty-three undergraduates at Indiana University participated in exchange for course credit.

Materials

Objects were taken from the Snodgrass and Vanderwart (1980) corpus of 260 objects. 240 of those objects have a single word label and all stimuli used in the following experiments were drawn from this set of 240. Note due to constraints on the materials, letter frequency is not controlled in these studies. The objects are well characterized and normed on a number of dimensions including visual complexity and WF, which are not correlated with one another (see Snodgrass & Vanderwart). The labels range in normative word frequency from 0 to 897 per million (Kucera & Francis, 1967). People are generally very accurate at naming the pictured objects. The highest rate of failure to provide a name for any single picture was 1.7% (Snodgrass & Vanderwart, 1980). Snodgrass and Yuditsky (1996) reported that 89% of the words are correctly identified by at least 90% of the participants (and 76% of the words are correctly identified by at least 95% of participants). For this free recall experiment it was critical that the word recalled by the participant match the object label provided by the corpus. Thus, the 60 objects used here were those whose label was correctly identified most often (all of which were correctly named by at least 95% of participants).

Design and procedure

Four 15-object lists were randomly created for each subject. Two study lists consisted of LF objects (1–4 per million) and two study lists consisted of HF objects (59–591 per million). Participants were instructed that they would study a series of words or objects one at a time, and that they should try to remember these stimuli for a future, memory test. Each object was presented for 1.5 s with no inter-stimulus interval. Following a 30 s. distractor task participants were instructed to recall as many items from the most recent list in any order they chose. This study-test cycle was repeated for each of the four lists in a randomly determined order.

Results and discussion

The results of this experiment extended the pure list normative frequency effect to objects: objects with HF labels were recalled reliably more often than objects with LF labels [.57 vs. .36, $t(32) = 7.13$]. Hence, normative object and word frequency affect free recall similarly, and a null effect of object frequency on recognition memory may not be attributed to the fact that object frequency does not affect episodic memory.[7] Obviously, this experiment does not answer the question of whether objects and their corresponding labels are experienced with similar frequency in the environment. However, the study does provide evidence that object frequency (defined as frequency of the corresponding label) behaves in a manner similar to word frequency.

Experiment 3b

In this study, we test the assumption that semantic information does not contribute to the mirror patterned word frequency effect by presenting participants with all combinations of studying words or objects and testing words or objects. In the two mixed conditions (i.e., study one type and test the other type), there is no overlap between the physical characteristics of the items studied and items tested. Semantic information must be used in these conditions in order to perform above chance. Given the large number of experiments suggesting that words and objects are associated with the same semantic codes in memory (see Nelson, 1979, for a review), performance should be above chance in the mixed conditions if participants store semantic information at study and if they probe with semantic information at test. Importantly, the EPEA hypothesis assumes that semantic information does not contribute to the word frequency effect (Malmberg & Nelson, 2003) and hence we do not expect a LF advantage in the mixed conditions. In addition, we assume that visual object features do not vary in frequency (or alternatively the frequency of visual object features is not correlated with the frequency of letters in the object's label). Thus, the only condition in which the EPEA hypothesis predicts an effect of prior normative frequency is when words are studied and words are tested.

[6] The HF advantage in recall is restricted to pure lists; studying a list of mixed frequency produces inconsistent results (e.g., Gillund & Shiffrin, 1984).

[7] Readers may be concerned that the benefit for HF objects is simply due to differences in the accuracy with which subjects name HF and LF objects. This is not the case. All objects used in this experiment are correctly named at least 95% of the time. In addition, normative frequency is uncorrelated with ability to generate a name (Snodgrass & Vanderwart, 1980), though people are faster to name HF objects than LF objects (Snodgrass & Yuditsky, 1996).

Method

Participants and materials

Two hundred and two undergraduate students at Indiana University participated in exchange for course credit. There were four between-subject conditions in this single item yes–no recognition experiment: Words were studied and words were tested (46 participants), objects were studied and objects were tested (64 participants), words were studied and the corresponding objects were tested (46 participants), or objects were studied and the corresponding words were tested (46 participants). Additional subjects were run in the object–object condition in order make sure that a small effect could be detected if it were present even if performance was very accurate due to the picture superiority effect.

For each subject, all 240 items from the Snodgrass and Vanderwart (1980) corpus (see above) were randomly assigned to be either a target or a foil. Each of the 120 targets was studied for 2.0 s with no inter-stimulus interval. The study list was followed by a 30-s distractor task consisting of mentally adding single digits. The test list consisted of 120 targets and 120 foils randomly intermixed. Due to the limited number of stimuli, we did not divide the stimuli into separate HF and LF groups. Instead, we used all stimuli and linear regression analyses to assess the ability of WF to predict the probability of calling an item "studied" across the four conditions.

Design and procedure

Participants were instructed that they would study a series of words or objects one at a time, and that they should try to remember these stimuli for a future, unspecified memory task. After the distractor task, participants were informed of the yes–no recognition task. When words were studied and objects were tested, participants were instructed to answer yes if "a word that names this object" was studied. When objects were studied and words were tested, participants were instructed to answer "yes" if a "picture of the word" was studied. In all conditions, examples were given. After indicating that they understood the instructions, participants proceeded to the self-paced recognition task.

Results and discussion

Average HRs and FARs and the results of paired *t*-tests are shown in Table 3. In each condition, the HR and FAR differ indicating that participants could discriminate between targets and foils, even when the decision was based on semantic information alone (i.e., when the visual form of encoding differed from the form at test).

Table 3
The average hit rate and false alarm rate for each condition in Experiment 3b

Studied–tested	Hit rate	False alarm rate	*t*-test
Word–Word	.699 (.010)	.330 (.017)	$t(79) = 18.358$, $SE = .020$, $p < .001$
Object–Object	.784 (.011)	.113 (.012)	$t(79) = 42.607$, $SE = .016$, $p < .001$
Object–Word	.676 (.015)	.348 (.019)	$t(79) = 13.055$, $SE = .025$, $p < .001$
Word–Object	.594 (.013)	.398 (.015)	$t(79) = 10.849$, $SE = .018$, $p < .001$

Standard errors of the mean are in parenthesis. In all cases, the hit rate is greater than the false alarm rate reflecting above chance performance.

Fig. 4 plots recognition memory as a function of normative frequency of the label separately for each of the four conditions. There were a total of 80 unique frequencies associated with the stimuli. Each point plots the probability that a stimulus (object or word) of a given normative frequency is called "old" given that the stimulus was a target or a foil. The black points are the HRs and the white points are the FARs, one point for each frequency category. The lines plotted in each graph are the result of linear regression analyses relating log normative frequency to HRs and FARs; the equation for the regression line is noted in each plot. To avoid calculating the log of zero, a constant (1) was added to the raw frequencies prior to the log transformation.

When words were studied and words were tested a mirror patterned word frequency effect was observed. A linear regression analysis revealed that increasing WF predicts a decreasing HR: ($t(1,79) = -2.61$, $p = .011$, $r = .284$, $SE = .008$) and an increasing FAR ($t(1,79) = 3.33$, $p = .001$, $r = .353$, $SE = .012$). In contrast, when objects were studied and tested, the word frequency of the label was not a reliable predictor of HRs ($t(1,79) = -1.21$, $p = .228$, $r = .136$, $SE = .009$) or FARs ($t(1,79) = -.298$, $p = .767$, $r = .034$, $SE = .009$). This is consistent with the hypothesis that the identification of uncommon words, but not objects, demands extra attention during study giving rise to the LF benefit in recognition memory.

Normative frequency was not a reliable predictor of endorsing an item as "studied" when words were studied and objects were tested for targets ($t < 1$, $r = .062$, $SE = .01$) or foils ($t < 1$, $r = .016$, $SE = .012$). When objects were studied and words were tested, WF did not predict FARs ($t < 1$, $r = .087$, $SE = .015$) though it did predict that HRs should increase with increasing WF ($t(1,79) = 2.00$, $p = .049$, $r = .221$, $SE = .012$), opposite the standard finding and opposite the pattern in the current experiment when words are studied and tested. These findings are consistent with the hypothesis that semantic information does not contribute to the word frequency effect (Criss & Shiffrin, 2004a; Malmberg & Nelson, 2003).

The above analyses report individual linear regressions for HR and FAR for each of the four conditions defined by the stimulus material at study and test. These individual analysis suggest an interaction in that WF meaningfully predicts P(old) only in the condition where words are studied and tested, as predicted by the EPEA hypothesis. This interaction was directly tested with a moderated linear regression approach including log WF, condition, and an interaction term as predictors (Aiken & West, 1991). Log WF was centered (a linear transformation setting the mean at 0) as is standard for continuous predictor variables (Aiken & West, 1991). We were specifically interested in the contrast between the word–word condition and the remaining conditions thus condition was treated as a dichotomous variable (word–word condition vs. other). The interaction term was the product of the two predictors. To assess the importance of the interaction term, we report both the interaction coefficient and the change in model fits for models with and without the interaction term. For

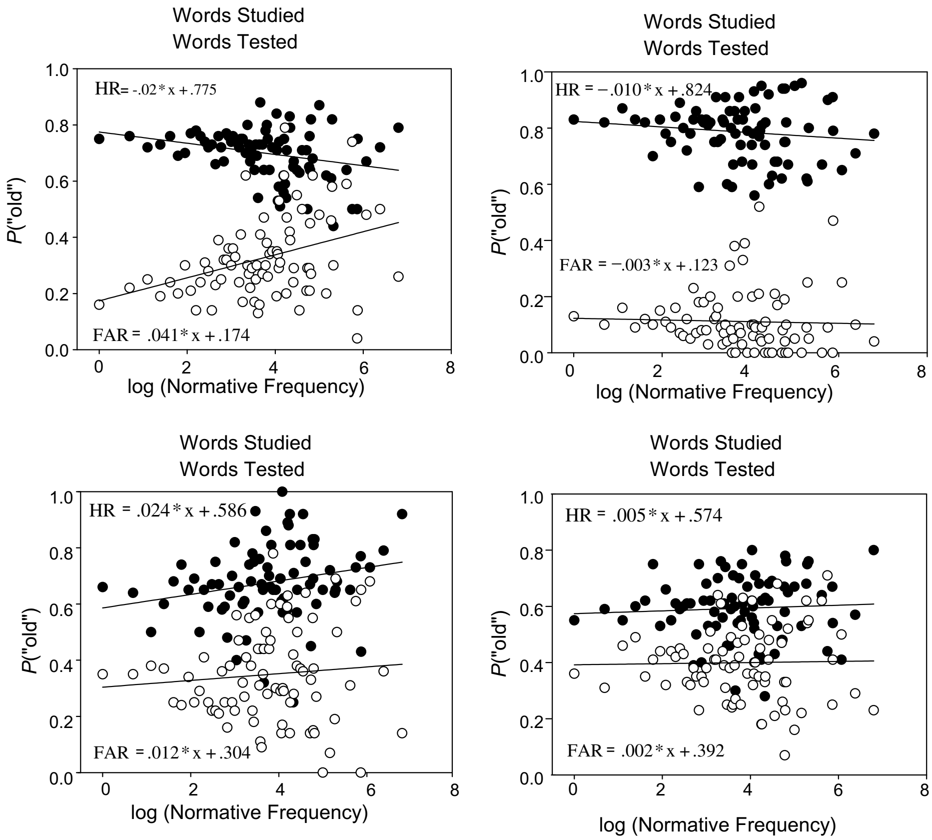

Fig. 4. The results of Experiment 3b. Targets are black circles and foils are white circles. Regression lines and equations are plotted for each condition.

hit rates, including the interaction term significantly improved the amount of variance accounted for by the model (change in R^2 = .012, $F(1,316)$ = 3.86, p = .050, b for the interaction term = .026, SE = .013). Likewise, for false alarm rates, including the interaction term in the model significantly improved the amount of variance accounted for (change in R^2 = .014, $F(1,316)$ = 4.47, p = .035, b for the interaction term = −.038, SE = .018). As suggested by Fig. 4 and the individual linear regression analyses, WF predicts P(old) when words are studied and tested but not when objects are studied, tested, or both.

The results of this study support predictions made by the EPEA hypothesis. We find an increasing HR with decreasing WF when words are studied and tested and thus orthographic/phonemic features are encoded in the memory trace and provided as a cue during the memory test. We propose that this results from the difficulty in using such features, predominant in LF words, to identify the word. That difficulty is accommodated by allocating extra resources during the early stage of encoding resulting in a more accurate and/or diagnostic memory trace. When objects are encoded there is no differential distribution of attention during identification and thus no benefit for LF targets. In mixed conditions where the visual features do not overlap at encoding and retrieval, performance is based on semantic features. We propose that semantic features do not contribute to the WFE because they are not differentially diagnostic for LF and HF words as are orthographic/phonemic features. The opposite pattern for HRs in the condition where objects were studied and words were tested is curious and deserves investigation in future studies.

General discussion

The goal of this manuscript is to evaluate the EPEA hypothesis. According to this hypothesis, the early stage of processing involves identification and the subsequent extraction of the meaning of a word. The more difficult a word is to identify, the more attention it requires during this early stage, and thus the stored memory trace is more accurate and/or diagnostic. In contrast, the late stage of processing involves elaboration, building associations, and other strategic processing. Critically, the EPEA hypothesis accounts for the LF HR advantage by acknowledging that LF words tend to be composed of distinct orthographic features that require more attention to process during the early stage. Simultaneously, the hypothesis assumes that semantic features are not differentially diagnostic as a function of normative frequency and thus the late stage, where these features garner attention, does not contribute to the WFE. Thus, the LF HR advantage is accumulated during the first 1 s. of encoding that makes up the early phase. The late phase, while obviously important for the overall accuracy of memory, does not contribute to the WFE.

In two experiments, we demonstrated a benefit for targets composed of uncommon letters when normative frequency was held constant. This benefit was present early in encoding and did not grow with additional study time. Further, this result is consistent across encoding conditions. When the encoding task is specified there is no benefit for LF targets if letter frequency is held constant. More variable encoding conditions (i.e., when there is no assigned task) provide variable results. Experiment 2 demonstrated no difference between HF and LF targets encoded with no specific task while Experiment 1 demonstrated a marginally significant LF HR advantage. The strategies used by participants when provided with no guidance at encoding deserve further attention that will inform and hopefully resolve this issue. Overall the data support the hypothesis that the HR is proportional to the difficulty in decoding the orthography in order to identify the word. Unusual orthography, as measured by letter frequency, leads to additional attention and/or effort during the early phase of encoding which improves episodic memory.

Desirable difficulties enhance memory

The idea of a desirable difficulty, introduced by Bjork (1994), is that factors that make learning arduous often lead to better long term learning and memory performance. According to the EPEA hypothesis, the WFE for recognition memory is another instance where the difficulties associated with identifying words comprised of unusual letters produces a desirable outcome. An avenue ripe for future research may be investigating the interaction between letter frequency and other forms of desirable difficulties during study (e.g., spaced vs. massed presentations).

Measures of orthography

Our measure of letter frequency is theoretically grounded in the REM framework for episodic memory (Shiffrin & Steyvers, 1997) but has not been investigated (to our knowledge) in the word recognition literature. Instead neighborhood density, and to a lesser extent bigram frequency, are the common measures of orthography in that literature. The impact of orthography on word identification has been widely studied. In contrast, just a few studies have examined the impact of these variables on recognition memory.

First consider neighborhood density, defined as the number of words formed by changing a single letter of a given word. For example, the neighbors of *hope* include *pope*, *rope*, *cope*, *home*, *hole*, *hose*, *hops*, *hone*, etc. While there is much debate, our reading suggests that neighborhood density has opposing effects on two tasks that measure word recognition: high density facilitates lexical decision but harms naming (e.g., Andrews, 1997; Norris, 2006; Perea & Rosa, 2000; Pollatsek, Perea, & Binder, 1999).[8] Heathcote, Ditton, and Mitchell (2006) orthogonally manipulated neighborhood density and word frequency in a recognition memory study. They found simultaneous mirror effects for neighborhood density when word frequency was held constant and for word frequency when neighborhood density was held constant. Noting that neighborhood density is typically correlated with bigram frequency, Heathcote et al. conducted further analyses on subsets of their data and report that the neighborhood density mirror pattern holds and cannot be fully explained by differences in bigram frequency. They computed letter frequency for their stimulus sets and it did not differ for the high and low density conditions, thus they did not analyze the impact of letter frequency on recognition memory.

A subset of neighborhood density as just defined is orthographic neighborhood size which is the number of items that share orthography and phonology with a given word. For example hope has an orthographic neighborhood size of 7 including *lope*, *mope*, *dope*, *cope*, *pope*, *rope*, *and nope* (note that *home*, *soap*, *hops*, etc. are not included in the neighborhood in this measure). Two recognition memory studies manipulated orthographic neighborhood size while holding word frequency constant (Cortese, Watson, Wang, & Fuggett, 2004; Glanc & Greene, 2007). Cortese et al. found a greater HR but not FAR for words with a small orthographic neighborhood. Glanc and Greene found a full mirror pattern. Glanc and Greene also found that an animacy orienting task eliminated the benefit for targets with a small neighborhood. In contrast, the current findings demonstrate a benefit for words with distinct letters across all orienting tasks. Therefore, we computed letter frequency estimates for Glanc and Greene's stimuli. There was no difference in letter frequency between low ($M = .0730$, $SD = .0211$) and high ($M = .0731$, $SD = .0219$) density words (collapsing across their consistent/inconsistent condition). The different pattern of data between Glanc and Greene and the results of Experiments 1 and 2 suggests that orthographic neighborhood density and letter frequency make separate contributions to episodic memory.

A limited number of studies have concluded that there is little or no impact of bigram frequency on word recognition memory (e.g., Andrews, 1992; Gernsbacher, 1984) and we found no studies that measured the impact of bigram frequency on recognition memory (see Heathcote et al. 2006 for some secondary analyses of bigram frequency). There are number of different ways to measure bigram frequency (e.g., whether one takes into account word length, frequency of the word in which the bigram appears, or position of the bigram within the word; whether the sum or the mean of individual bigram frequencies is used, etc.) and the potential impact of each of these different measurements on word recognition and episodic memory deserve further exploration.

Together with the current studies, this set of research indicates that the impact of orthography on episodic memory should be carefully examined in future research. Questions for future research include: What are the differences between these four manipulations (letter frequency, neighborhood density, orthographic neighborhood size, and bigram frequency) in terms of structure of the resulting

[8] Additional studies have examined the contribution of the frequency of a word's neighbors (e.g., Sears, Campbell, & Lupker, 2006), the role of neighbors that share phonology vs. neighbors that share orthography (e.g., Adelman & Brown, 2007; Mulatti, Reynolds, & Besner, 2006), and the early role of meaning in word recognition (e.g., Pexman, Hino, & Lupker, 2004).

stimuli? What are the combined and independent contributions of each of these measures of orthography on episodic memory performance and on word recognition?

The role of semantic features on the WFE

Experiment 3 demonstrates a null mirror effect when objects are studied or tested but an intact mirror effect when words are studied (without any specific task) and words are tested. Thus, when we can be relatively certain that the basis for the recognition decision did not involve a comparison of the physical features of the test stimulus to the contents of memory, no word frequency effect was observed. This supports the hypothesis that semantic information, which is the focus of attention during the late phase, is not differentially diagnostic for word of differing normative frequency. Only when the orthographic features of the words are stored in the episodic memory trace and used in the retrieval cue do we see a LF HR advantage.

The sensory/semantic model, developed to account for the object superiority effect, claims that multiple different types of features may be stored in memory, depending on attention and the cue presented for encoding (e.g., Nelson, 1979; Nelson & Brooks, 1973; Nelson et al., 1977). Specifically, objects result in the storage of visual features with a direct pathway to meaning while words provide an orthography-to-meaning pathway and an orthography-to-phonology-to-meaning pathway. We assume that the orthography and possibly phonology features lead to the LF HR advantage. When an object is presented as an encoding or retrieval cue these features are not directly available and are not part of the cue (unless specific instructions direct otherwise, cf., Criss & Shiffrin, 2004a). The finding of a WFE for the word–word condition and no other conditions supports the predictions of the sensory/semantic model.

An alternative explanation for the findings of Experiment 3 is differences in stability of representation for objects and words.[9] Words are perceived as whole units but objects may have more variable representations due to the fact that the same conceptual object can be represented by many different viewpoints and different visual aspects. If the prior frequency effect is a function of the number of different contexts in which an item has appeared (i.e., Dennis & Humphreys, 2001 but see Criss & Shiffrin, 2004b) then a given instance of an object may not contact other instances of the same object due to this encoding variability (cf., McClelland & Chappell, 1998), thus eliminating our ability to measure a prior frequency effect for object stimuli. This seems a reasonable explanation for the lack of a prior frequency effect when objects are studied and/or tested. In fact, this proposal has been used to account for better performance for LF (more stable representation) than HF (more variable representation) words (Criss & McClelland, 2006; McClelland & Chappell, 1998).

However, this very same explanation predicts better performance for the class with a stable representation (words) than a variable representation (objects). This prediction is clearly at odds with the classic picture superiority effect (Nelson, Reed, & Walling, 1976; Shepard, 1967) demonstrating better memory for objects than words and replicated here. No doubt there are many possible alternatives to explain why we find a full mirror pattern prior frequency effect for the word–word condition but no conditions involving an object at study and/or test. However, we can think of no alternative that accounts for the object superiority effect, the effect of letter frequency, and the time course of the WFE and thus prefer the EPEA hypothesis.

False alarm rates

The focus of this EPEA hypothesis and this paper is the process of encoding information into memory. As such, we have focused almost entirely on HRs. Our hypotheses (Criss & Shiffrin, 2004a; Malmberg & Nelson, 2003) make no claims about the nature or the underlying cause for the FAR portion of the word frequency mirror effect. In the current studies, we found a higher FAR for HF than LF words when words were studied and tested but not when objects were studied and/or tested. In addition, we found a higher FAR for words composed of typical than distinct letters.

Together these findings suggest that letter frequency also contributes to the FAR portion of the word frequency effect. This is consistent with the REM model (Shiffrin & Steyvers, 1997). According to REM, HF words tend to be composed of common features (letters in this case) that are shared with many other words. Thus, HF foils tend to match features of many other words by chance (other HF words in particular) and thus feel rather familiar and will be called "studied" on this basis. LF words are composed of less common/more diagnostic features. LF foils tend to match fewer features by chance resulting in a lower FAR. Critically, the amount of evidence provided by matching a feature stored in episodic memory is mediated by the diagnosticity of that feature value: the more diagnostic the feature value (i.e., the less common the feature) the more evidence it provides which results in a LF HR advantage.

The EPEA hypothesis has not yet been incorporated into REM. The necessary steps would include defining a subset of item features that do not differ in diagnosticity as a function of normative WF (i.e., semantic features) and to define a time course for encoding with orthographic features (that do differ in diagnosticity) stored during an early phase and semantic features stored during a late phase.

The familiarity based explanation for the higher FAR for HF foils is consistent with many other models of recognition memory, though the underlying cause of the increased familiarity varies between models. Some assume the higher familiarity arises due to the number of prior contexts in which HF words have appeared (Dennis & Humphreys, 2001; Reder et al., 2000), due to encoding variability (McClelland & Chappell, 1998), among other possibilities. All of the theories just described claim that FAR is based on conceptual information (rather than the form of the word itself) and thus would seem to predict a higher FAR for HF than LF words even when objects are studied and/or tested, which is inconsistent with our findings.

[9] We thank Simon Dennis for this suggestion.

A comparison of SAC and the EPEA hypothesis

As noted earlier, Malmberg and Nelson (2003) found poorer performance for words paired with a LF than with a HF word during encoding, consistent with the elevated-attention hypothesis. Diana and Reder (2006) replicated and extended the finding to pictures, demonstrating the same decline in performance for pictures paired with LF words relative to those paired with HF words. Diana and Reder agree that LF words require more attention during the early stage of encoding. However, they attribute the difficulty of encoding to binding the word with the current context rather than identification of the word itself. Further, they maintain that the LF HR advantage is due to better recollection of LF than HF words during retrieval. A recent augmentation of the Source of Activation Confusion (SAC) model implements an inverse relationship between item familiarity and the amount of working memory resources required to encode the stimulus (Reder, Paynter, Diana, Ngiam, & Dickinson, 2008).

The SAC model assumes that LF words are linked to fewer episodes than are HF words (Reder et al., 2000). Despite the lower pre-experimental strength (i.e., familiarity) of the LF words, they are more likely to be recollected in a particular episode because their strength spreads to fewer connected episodes than a HF word. Thus according to SAC, a LF word is more likely to be recollected from a given list than a HF word from the same list. This retrieval based explanation accounts for the WFE especially as measured by the remember/know paradigm. A recent addition to the SAC model acknowledges the role of encoding and the differential attentional resources allocated to LF and HF words at study (Reder et al., 2008). They assume that LF words are more difficult to encode and therefore require more attention than HF words.

At first glance, the new SAC assumption and the hypotheses presented in this paper and its predecessors (i.e., Criss & Shiffrin, 2004a; Malmberg & Nelson, 2003) are very similar. Indeed, we agree that LF words are more difficult to encode and that this results in LF words absorbing more attention during encoding than HF words. However, there are a number of points, some more subtle than others, where the two theories diverge. For purposes of clarity and to provide a roadmap for future research, we outline these differences in some detail.

The EPEA hypothesis attributes the difficulty in encoding LF words to the identification stage and specifically to the unusual orthography (and possibly phonology) of LF words. SAC attributes the difficultly to familiarity—a less familiar stimulus is more difficult to encode. The SAC explanation is more general but it does not provide a compelling answer to *why* less familiar stimuli should be more difficult to encode. Critically, this seems to predict a prior frequency effect for objects—less familiar objects should be more difficult to encode and thus better recollected—which we do not find here. Second, our explanation incorporates a time course for this effect. The EPEA hypothesis claims that any extra attention required by LF words is resolved in the first 1 s. of encoding. Attention during the later phase of controlled processing is not differentially distributed to LF and HF words. This is supported by the presence of a LF HR advantage by approximately 1 s of study that does not grow in size with additional study time (Criss & McClelland, 2006; Criss & Shiffrin, 2004a; Estes & Maddox, 2002; Malmberg & Nelson, 2003). SAC makes no specific claims about the time course of encoding. Third, we suggest that the attention received by LF words during the early phase results in a more diagnostic memory trace while SAC assumes that it is required to bind the concept and episode nodes. Finally, in SAC the retrieval advantage for LF words is strong enough to produce the LF HR advantage *despite* an encoding disadvantage. In our view, the extra attentional resources at encoding and the greater familiarity during retrieval both result from the same source: the uncommon features from which LF words are composed.

Thus, the SAC view borrows many the assumptions of EPEA hypothesis (i.e., LF words require additional attention due to difficulty in encoding), but we disagree on many of the finer points. Of course, the devil is in the details and even models that share nearly all high-level assumptions but differ in the details of implementing these assumptions often make different qualitative and quantitative predictions (cf., Criss & McClelland, 2006). For this reason, it seems worthwhile to explore these points of digression in future research.

Conclusions

These experiments explored several assumptions of the EPEA account of the word frequency effect for recognition memory. First, we considered the hypotheses that the early encoding phase, in which words are identified, is the source of the word frequency effect. Because LF words are composed of unusual, and therefore difficult to encode, features, they require more attention during this stage. We found an advantage for targets composed of distinctive letters and a noisy (Experiment 1) or null (Experiment 2) word frequency effect when letter frequency was held constant. The current study adds to a number of recent studies emphasizing the role of orthography in recognition memory (Cortese et al., 2004; Glanc & Greene, 2007; Heathcote, et al. 2006). Next, we considered the late phase of encoding, in which meaning is elaborated and connected to the participants' semantic network. We explored the hypothesis that semantic information does not contribute to the word frequency effect. We found a mirror patterned normative frequency effect when words were studied and tested. When study, test, or both involved objects, the mirror pattern was disrupted and normative frequency had little or no effect on recognition memory.

Much of the recognition memory literature has focused on distinguishing between memory models based on the nature of retrieval from memory (e.g., Yonelinas, 2002). However, the EPEA hypothesis adopts a different approach to understanding prior frequency effects by assuming that normative frequency influences encoding. The current study adds to the growing consensus that understanding encoding operations is central to understanding the WFE in particular and recognition memory more generally (cf., Criss, 2006; Criss & Shiffrin, 2004a, 2004c; Hirshman, Fisher, Henthorn, Arndt, & Passannante, 2002; Malmberg, Holden, & Shiffrin, 2004).

Acknowledgments

Portions of this research were supported by National Institutes of Mental Health Training Grant MH019102. A subset of this research were presented at the Thomas O. Nelson Memorial Symposium in Toronto (November, 2005). The authors thank Alecia Santuzzi for assistance with the linear regression analyses. The authors also thank Bob Greene, Bill Hockley, and two anonymous reviewers.

References

Adelman, J. S., & Brown, G. D. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14*, 455–459.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Andrews, S. (1992). Neighbourhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning Memory and Cognition, 18*, 234–254.

Andrews, S. (1997). The role of orthographic similarity in lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*, 439–461.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed systems and its control processes. In W. E. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 742–775). New York: Academic.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium*. Philadelphia, PA: University of Pennsylvania.

Balakrishnan, J. D., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 615–633.

Besner, D., & McCann, R. S. (1987). Word frequency effects and pattern distortion in visual word identification and production: An examination of four classes of models. In M. Coltheart (Ed.), *Attention and performance. XII: The psychology of reading* (pp. 201–219). New Jersey: Erlbaum.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: The MIT Press.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 813–840.

Brown, J. (1976). *Recall and recognition*. New York: John Wiley Sons.

Cortese, M. J., Watson, J. M., Wang, J., & Fugett, A. (2004). Relating distinctive orthographic and phonological processes to episodic memory performance. *Memory & Cognition, 32*, 632–639.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language, 55*, 461–478.

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.

Criss, A. H., & Shiffrin, R. M. (2004a). Interactions between study task, study time, and the low frequency HR advantage in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 778–786.

Criss, A. H., & Shiffrin, R. M. (2004b). Context noise and item noise jointly determine recognition memory: A comment on Dennis & Humphreys (2001). *Psychological Review, 111*, 800–807.

Criss, A. H., & Shiffrin, R. M. (2004c). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition, 32*, 1284–1297.

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination and representation in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1199–1212.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452–478.

Diana, R. A., & Reder, L. M. (2006). The low frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 805–815.

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1075–1095.

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

Forster, K. I. F., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*, 256–281.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.

Glanc, G. A., & Greene, R. L. (2007). Orthographic neighborhood size effects in recognition memory. *Memory & Cognition, 35*, 365–371.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 12*, 8–20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Theory and data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 5–16.

Guttentag, R. E., & Carrol, D. (1997). Recollection-based recognition: Word-frequency effects. *Journal of Memory and Language, 37*, 502–516.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1511–1521.

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

Hintzman, D. L. (1994). On explaining the mirror effect. *Memory & Cognition, 20*, 201–205.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative explanations of false recognition: The case of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1306–1323.

Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word frequency mirror effect. *Journal of Memory and Language, 47*, 499–516.

Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory, but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning Memory and Cognition, 17*, 507–513.

Hirshman, E., Trembath, D., & Mulligan, N. (1994). Theoretical implications of the mnemonic benefits of perceptual interference. *Journal of Experimental Psychology: Learning Memory and Cognition, 20*, 608–620.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480–517.

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.

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

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.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. (1976). Depth of processing, recognition, and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 75–102). NY: Wiley.

Malmberg, K. J., Holden, J., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 319–331.

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.

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

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

Mandler, G., Goodman, G. O., & Wilkes-Gibb, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition, 10*, 33–42.

Martin, A., & Wiggs, C. L. (1997). Modulation of human medial temporal lobe activity by form, meaning, and experience. *Hippocampus, 7*, 587–593.

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.

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture–word interference paradigm. *Journal of Experimental Psychology: General, 132*, 228–258.

Mulatti, C., Reynolds, M., & Besner, D. (2006). Neighbourhood effects in visual word recognition: New findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 799–810.

Mulligan, N. W. (2000). Perceptual interference at encoding enhances item-specific encoding and disrupts relational encoding: Evidence from multiple recall tests. *Memory & Cognition, 28*, 539–546.

Murdock, B. (2003). The mirror effect and the spacing effect. *Psychonomic Bulletin & Review, 10*, 570–588.

Naveh-Benjamin, M., Craik, F. I. M., Guez, Y., & Dori, H. (1998). The effects of divided attention on encoding and retrieval processes in human memory: Further support for an asymmetry. *Journal of Experimental Psychology, 24*, 1091–1104.

Naveh-Benjamin, M., & Guez, Y. (2000). The effects of divided attention on encoding and retrieval processes: Assessment of attentional costs and a componential analysis. *Journal of Experimental Psychology, 26*, 1461–1482.

Nelson, D. L. (1979). Remembering pictures and words: Appearance significance and name. In L. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory* (pp. 45–76). New Jersey: L. Erlbaum & Associates.

Nelson, D. L., & Brooks, D. H. (1973). Functional independence of pictures and their verbal memory codes. *Journal of Experimental Psychology, 98*, 44–48.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 125–173). New York: Academic Press.

Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 485–497.

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 523–528.

Norris, D. (2006). The Bayesian Reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*, 327–357.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

Perea, M., & Rosa, E. (2000). Repetition and form priming interact with neighborhood density at a short stimulus-onset asynchrony. *Psychonomic Bulletin & Review, 7*, 668–677.

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 1252–1270.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343*, 263–266.

Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1142–1158.

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.

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2008). Experience is a double-edged sword: A computational model of the encoding/retrieval tradeoff with familiarity. In B. Ross & A. S. Benjamin (Eds.), *The psychology of learning and motivation* (pp. 271–312). Academic Press.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 1–17.

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

Sears, C. R., Campbell, C. R., & Lupker, S. J. (2006). Is there a neighborhood frequency effect in English? Evidence from reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1040–1062.

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition, 21*, 34–40.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and 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.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592–604.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, and Computers, 28*, 516–536.

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

Underwood, B. J. (1969). Attributes of memory. *Psychological Review, 76*, 559–573.

Westerman, D. L., & Greene, R. L. (1997). The effects of visual masking on recognition: Similarities to the generation effect. *Journal of Memory and Language, 37*, 584–596.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517.

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