Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 4, 778–786

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.4.778

Interactions Between Study Task, Study Time, and the Low-Frequency Hit Rate Advantage in Recognition Memory

Amy H. Criss and Richard M. Shiffrin
Indiana University

In studies of episodic recognition memory, low-frequency words (LF) have higher hit rates (HR) and lower false alarm rates (FAR) than do high-frequency words (HF), which is known as the mirror pattern. A few findings have suggested that requiring a task at study may reduce or eliminate the LF-HR advantage without altering the LF-FAR effect. Other studies have suggested that the size of the LF-HR advantage interacts with study time. To explore such findings more thoroughly and relate them to theory, the authors conducted 5 experiments, varying study time and study task. The full mirror pattern was found only in 2 cases: the standard condition requiring study for a later memory test and a condition requiring a judgment about unusual letters. The authors explain their findings in terms of the encoding of distinctive features and discuss the implications for current theories of recognition memory and the word frequency effect.

Studies of episodic recognition memory typically demonstrate a mirror pattern for words varying in environmental frequency (e.g., Glanzer & Adams, 1985). Low-frequency words (LF) have a higher hit rate (HR) and a lower false alarm rate (FAR) than do high-frequency words (HF). HR is defined as the probability of calling a studied word old and FAR is defined as the probability of calling a nonstudied word old. This reversal in the probability of saying old to words varying in frequency is the mirror patterned word frequency effect (WFE). The WFE is so robust that it has been deemed one of the regularities of recognition memory (Glanzer, Adams, Iverson, & Kim, 1993) and has been the source of much empirical and theoretical work.

Accounting for the WFE has been an important benchmark for recent theoretical development (Dennis & Humphreys, 2001; Estes & Maddox, 2002; Glanzer & Adams, 1990; McClelland & Chappell, 1999; Murdock, 2003; Reder et al., 2000; Shiffrin & Steyvers, 1997; Sikström, 2001). One class of models, the global matching models, assume old–new recognition decisions are based on the overall match between a test item and the episodic memory traces resulting from the study list. Early versions of such theories required the assumption that different criteria are used for words of different frequency (e.g., Gillund & Shiffrin, 1984; Hintzman, 1994). More recent models attempt to find principled reasons for the mirror effect and for distortions of that effect. Collecting evidence concerning those conditions that produce the mirror effect and those that distort that pattern was the prime motivation for the present study.[1]

Effects of Study Task

Although the FAR portion of the WFE is rarely disrupted (cf. Malmberg, Holden, & Shiffrin, 2004), several studies have shown disruption of the HR portion of the WFE. For example, Alzheimer's disease patients and healthy participants given Midazolam, a drug that induces temporary anterograde amnesia, show reduction of the LF-HR advantage while maintaining the standard FAR pattern (Balota, Burgess, Cortese, & Adams, 2002; Hirshman, Fisher, Henthorn, Arndt, & Passannante, 2002). In another example, Stretch and Wixted (1998) selectively strengthened HF targets by presenting them several times during study while presenting LF targets once. This within-list strength manipulation produced an HF-HR advantage. Other studies limited resources at study and/or test and produced mixed results regarding the presence or absence of an LF-HR advantage (Balota et al., 2002; Hintzman, Caulton, & Curran, 1994; Joordens & Hockley, 2000; Kim & Glanzer, 1993).

Of particular importance for this article, a few studies have reduced or eliminated the HR portion of the WFE when participants were instructed to perform certain tasks at study (i.e., tasks other than the standard instructions to "remember these words for a later test"). Hirshman and Arndt (1997), for example, found similar HRs for LF and HF for concreteness judgments but not commonness judgments. Implications for theory are difficult to draw, however, because these studies used unlimited study time, a 24-hr retention interval, and a between-subjects manipulation of encoding task. Other evidence comes from Guttentag and Carroll (1997, Experiment 2) who found a smaller LF-HR advantage following pleasantness ratings compared with standard instructions. In another example, Mandler, Goodman, and Wilkes-Gibbs

This research was supported by National Institute of Mental Health Grant 12717 to Richard M. Shiffrin. We thank Ken Malmberg and E. J. Wagenmakers for many helpful discussions.

Correspondence concerning this article should be addressed to Amy H. Criss or Richard M. Shiffrin, Department of Psychology, Indiana University, Bloomington, IN 47405. E-mail: acriss@indiana.edu or shiffrin@indiana.edu

[1] Note that other variables may produce mirror patterns. Stretch and Wixted (1998) argued for two types of mirror effects, one based on strength (e.g., repetitions of study items) and one based on stimulus properties such as normative word frequency or concreteness. We focused solely on stimulus-based mirror patterns and the word-frequency mirror pattern in particular (see Joordens & Hockley, 2000, for a proposal of why other stimulus-based mirror patterns such as concreteness are less reliable).

(1982, Experiment 2) instructed participants to describe semantic or physical aspects of the study words and found at most a small LF-HR advantage (2%). (Unfortunately the relevant statistics are not reported and interpretations are made difficult by the unusually long study time of 20 s per item). Finally, Hilford, Glanzer, and Kim (1997) asked participants either to judge whether the study item was common or to count the number of letters in the study word. The ordering of the forced choice conditions generally conforms to the mirror pattern with one important exception. The probability of choosing the HF alternative in the null condition in which both items were studied was close to .50. Although these studies obviously differ in too many ways to draw strong conclusions, they suggest that answering a question about each study item tends to reduce or eliminate the HR portion of the WFE.

A noted exception to this suggested pattern occurs when lexical decision is used as the study task. Such studies generally find an LF-HR advantage (Hilford et al., 1997; Hoshino, 1991; Joordens & Hockley, 2000; Mandler et al., 1982; Rao & Proctor, 1984). Lexical decision differs from the tasks mentioned above on many factors. For one, nonwords are included on the study (and sometimes test) list, and studies have shown that list composition affects the word frequency (WF) pattern (Dorfman & Glanzer, 1988; Malmberg & Murnane, 2002). In addition, lexical decision is usually conducted under speed stress or with instructions to respond as quickly as possible without compromising accuracy. The usual results are unequal error rates and unequal reaction times (and thus study times) between LF and HF words. In consideration of these complications, the present study focuses on study tasks other than lexical decision.

Effects of Study Time

The relatively few studies of the effect of study time on the WFE have produced mixed results. Some studies have shown no change in the WFE with increases in study time (Estes & Maddox, 2002, for study durations of 400 vs. 1200 ms; Hirshman & Palij, 1992, for study durations of 800; 1,000; 1,200; and 2,500 ms). However, Kim and Glanzer (1993) found a decrease in the HR distributions and increase in the FAR distributions for both HF and LF words as study time was reduced (from 2 to 1 s). Attempting to reconcile these findings, Malmberg and Nelson (2003) found that the magnitude of the LF-HR advantage grew over about the first second of study and then remained stable for longer study times. To explain this result, they proposed an early phase elevated-attention model with two stages of study. During the first phase, the study word is read and its meaning retrieved from memory. LF words attract more attention during the first phase and are therefore stored more accurately or distinctively. According to their hypothesis, the entirety of the LF-HR advantage is due to the additional attentional resources accruing to LF words during this early phase. Control processes, such as imagery, sentence formation, or rote rehearsal, take place during the second phase and are assumed to be equivalent for HF and LF words.

Models of the WFE often assume a difference in some underlying property of LF and HF words (such as feature frequency in the retrieving effectively from memory [REM] model of Shiffrin & Steyvers, 1997). Such assumptions, in conjunction with a decision rule that combines information about the probability of the test item being old and the probability of the item being new, produce the mirror pattern (e.g., Dennis & Humphreys, 2001; Glanzer & Adams, 1990; Shiffrin & Steyvers, 1997; Sikström, 2001). When performance decreases because of a list-wide manipulation, such as a decrease in study time or shallow encoding processes, these models predict a simultaneous movement of HRs and FARs toward the midpoint, called concentering: HRs decrease and FARs increase. In addition, these models can predict that a within-list strength manipulation will change the ordering of HRs without changing the ordering of FARs (e.g., repetitions of HF but not LF words produce HF-HR > LF-HR; Sikström, 2001; Stretch & Wixted, 1998). However, these models do not account for interactions with study task or study time when strengthening, or other manipulations are applied equally to HF and LF words. These are the manipulations explored in this article, and the results are used to refine theories of the WFE.

Dual process models also have been used to explain the WFE and the associated pattern of hits and false alarms (e.g., Mandler, 1980). A newer subset of dual process models attribute the false alarm advantage of LF words to a familiarity process, and the hit advantage of LF words to recollection (Joordens & Hockley, 2000; Reder et al., 2000; Yonelinas, 2002). The basic idea is that HF words have higher familiarity and thus have higher FARs. However, LF words are easier to recollect, and this overcomes the initial HF benefit to produce higher HRs for LF words. Although LF words may perhaps be easier to recollect, it is an open question how often and under what circumstances such processes are used in recognition. For example, Yonelinas (2002) proposed that semantic tasks increase recollection compared with perceptual tasks. In addition, others have proposed that conditions reducing recollection will reduce or eliminate the LF-HR advantage (Balota, Burgess, Cortese, & Adams, 2002; Hirshman et al., 2002; Joordens & Hockley, 2000). Our investigations of the relation of the WFE and study task therefore shed light on the use of recollection in recognition and the dual process account of the WFE.

General Method

The five experiments have similar designs; the differences largely lie in the tasks given at study. The basic design is given below and the exceptions are described separately for each experiment.

Stimuli and Design

The word pool consisted of 262 LF words, occurring between 1 and 10 per million, and 262 HF words, occurring 50 per million or more (Kucera & Francis, 1967). Words were assigned to each condition randomly and anew for each participant.

Word frequency (HF vs. LF) and study time (0.5, 1, or 3 s) were manipulated within participant, within list. The study list consisted of 120 words, half LF and half HF presented singly with a 150 ms interstimulus interval (ISI). Of the 60 words from each frequency, an equal number of words were assigned to each study time. All participants were informed that an unspecified memory test would follow. After 30 s of math problems, 240 items were presented singly for a yes–no recognition memory test. The test items consisted of all 120 targets along with 60 HF and 60 LF foils, all randomly intermixed. An alpha level of .05 was used for all statistical tests.

Experiment 1

Experiment 1 set the stage by replicating findings (e.g., Hirshman & Arndt, 1997) that have shown that certain study tasks cause a reduction or elimination of the LF-HR advantage. This study provides stricter control of study time than previous studies.

Method

A total of 77 Indiana University students participated as part of a course requirement. The general methods apply with two amendments. First, six primacy trials began the study list and these words were not tested. Second, half of the participants were randomly assigned to the *no-task* condition ($N = 38$), corresponding to typical experiments in which participants are simply told to study the items for a later memory test. The other half were assigned to the *concreteness* condition ($N = 39$) in which they answered *yes* or *no* to the following question during study "Does this word represent something you can see, hear, taste, smell, or feel?" (following Hirshman & Arndt, 1997).

Results

Table 1 gives the FARs. A 2 × 2 mixed design analysis of variance (ANOVA; word frequency and study condition) confirmed the typical finding of a greater FAR for HF words, $F(1, 75) = 121.92$, $p < .001$, $MSE = .01$. No other main effects or interactions were significant (all Fs < 1).

Figure 1A gives the HRs. A 2 × 3 × 2 mixed design ANOVA (word frequency and study time as within-subject variables and study condition as the between-subjects variable) confirmed: The LF-HR was greater than the HF-HR, $F(1, 75) = 22.02$, $p < .001$, $MSE = .02$; performance increased with study time, $F(2, 150) = 35.82$, $p < .001$, $MSE = .01$; and the concreteness participants had higher HRs than did the no-task participants, $F(1, 75) = 4.56$, $p = .036$, $MSE = .10$. There was an interaction between study time and study condition, $F(2, 150) = 4.39$, $p < .014$, $MSE = .01$, due to HRs increasing monotonically with study time for the concreteness condition but not increasing between 500 ms and 1 s in the no-task condition. This interaction may seem curious, but the failure to see an increase in HRs with relatively short study times is more or less consistent with findings of Malmberg and Nelson (2003), who found no increase from 250 ms to 1 s for HF words and Estes and Maddox (2002), who found no increase from 400 to 1,200 ms for HF words. Of greatest importance, there was a WF by study condition interaction due to the marked reduction of the LF-HR advantage in the concreteness condition, $F(1, 75) = 4.03$, $p < .048$, $MSE = .02$, replicating Hirshman and Arndt (1997). No other main effects or interactions approached significance (all Fs < 1).

Experiment 2

In Experiment 1, the tasks used at study varied between participants. Perhaps the participants in the concreteness condition were

Table 1
False Alarm Rates for Low-Frequency (LF) and High-Frequency (HF) Words in Each Experiment

Experiment	LF	HF
Experiment 1 (concreteness)	.194	.329
Experiment 1 (no task)	.189	.314
Experiment 2	.253	.338
Experiment 3	.225	.318
Experiment 4	.172	.305
Experiment 5	.237	.333

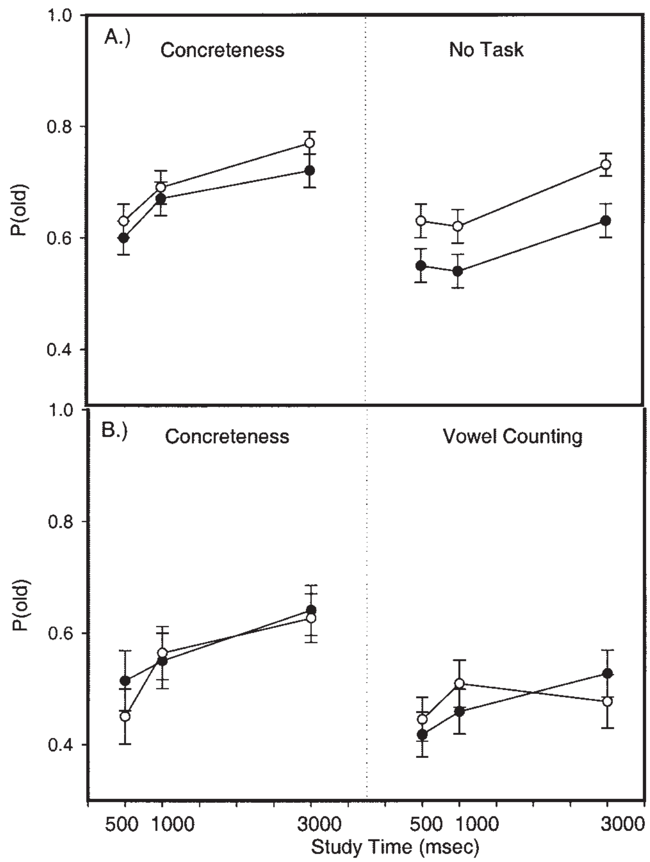

Figure 1. Hit rates (HRs) for low-frequency (LF) and high-frequency (HF) words as a function of study time and study task. Panel A shows HRs for Experiment 1, and Panel B shows HRs for Experiment 2. HF is indicated by a black circle; LF is indicated by a white circle. P = probability of calling the test item "old."

able to limit the features used at test to those features thought to be relevant, namely those that were emphasized by the concreteness task. If such semantic features do not differ for LF and HF words, this strategy could eliminate the LF-HR advantage. In the following experiment, we attempted to eliminate this strategy by mixing the concreteness task with a task that focused on lower level features.

Method

A total of 22 Indiana University students participated as part of a course requirement. The general methods apply with two amendments. First, six primacy trials began the study list (half were vowel counting and half were concreteness), and these words were not tested. Second, on half of the study trials, participants answered the concreteness question from Experiment 1. On the other half they answered *yes* or *no* to the following question that we refer to as *vowel counting*: "Does this word have more consonants than vowels?" The tasks were presented in alternating blocks of six trials of one task followed by six trials of the other task. A different background screen color was used for each task as an additional cue for the participant.

Results

Again, FARs to HF words were greater than LF words, $t(21) = 3.67$, $p < .001$, $SEM = .02$, as can be seen in Table 1. A 2 × 3 ×

2 repeated measures ANOVA (word frequency, study time, and task) was conducted on the HRs whose values are shown in Figure 1B. The concreteness task produced greater HRs than the vowel counting task, $F(1, 21) = 18.43$, $p < .001$, $MSE = .03$, and HRs improved with study time, $F(2, 42) = 15.83$, $p < .001$, $MSE = .02$. No other main effects or interactions approached significance (all Fs < 1.54). Notably, the LF-HR and HF-HR did not differ even when the concreteness and vowel counting tasks were mixed within list, strengthening the findings from Experiment 1.

Experiment 3

In Experiment 2, the study tasks alternated in short blocks. Experiment 3 strengthens the findings by alternating tasks randomly on a trial-by-trial basis and by mixing four study tasks.

Method

A total of 31 Indiana University students participated as part of a course requirement. The general methods apply with two amendments. First, 12 primacy trials began the study list (an equal number of each task), and these words were not tested. Second, an equal number of randomly selected words from the WF by study time combinations were assigned to each of four tasks. The four tasks were (a) the concreteness task, (b) an *animacy* task ("Does this word represent something that is living?"), (c) a *pleasantness* task ("Do you find this word pleasant?"), and (d) a *frequency* task ("Do you frequently encounter this word?"). A different background screen color was used for each task, and the tasks were randomly intermixed.

Results

As in the previous experiments, FARs to HF words were greater than to LF words, $t(30) = 4.89$, $p < .001$, $SEM = .02$, as can be seen in Table 1. HRs are presented in Figure 2A. A $2 \times 3 \times 4$ repeated measures ANOVA (word frequency, study time, and task) was conducted on these HRs. The study tasks led to varying levels of performance, $F(3, 90) = 10.30$, $p < .001$, $MSE = .03$. Overall, HRs increased with study time, $F(2, 60) = 65.36$, $p < .001$, $MSE = .07$, and there was a WF by study time interaction, $F(2, 60) = 3.14$, $p < .050$, $MSE = .04$. Observation of Figure 2A indicates that this interaction is due to the greater slope for the increase in performance for LF than HF words. No other main effects or interactions approached significance (all Fs < 1.28). Replicating the patterns of findings in the other studies, we saw no reliable LF-HR advantage when subjects were instructed to perform a task at study.

Experiment 4

Although Experiments 2 and 3 showed no reliable frequency differences in HRs when participants engaged in some study task, interpretation was difficult because these studies did not include a condition that reliably produced differences (e.g., instructions to simply study for a later memory test). Experiment 4 therefore included the no-task condition. For additional data on a low-level task, the vowel counting task was included.

Method

A total of 29 Indiana University students participated as part of a course requirement. The design was identical to Experiment 3 with one exception.

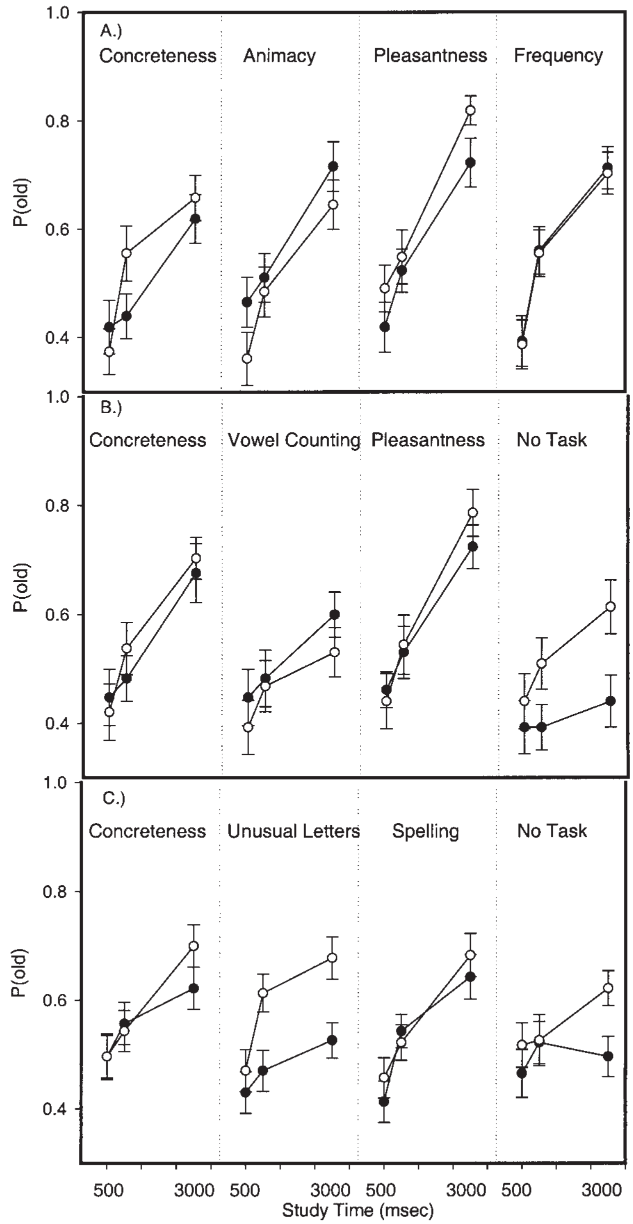

Figure 2. Hit rates (HRs) for low-frequency (LF) and high-frequency (HF) words as a function of study time and study task. Panel A shows HRs for Experiment 3, Panel B shows HRs for Experiment 4, and Panel C shows HRs for Experiment 5. HF is indicated by a black circle; LF is indicated by a white circle. P = the probability of calling the test item "old."

The study tasks were (a) concreteness, (b) pleasantness, (c) vowel counting, and (d) no task.

Results

Once more, FARs to HF words were greater than LF words, $t(28) = 6.63$, $p < .001$, $SEM = .02$. The values are reported in Table 1.

A $2 \times 3 \times 4$ repeated measures ANOVA (word frequency, study time, and condition) was conducted on the HRs presented in Figure 2B. Some study tasks resulted in greater HRs than others, $F(3, 84) = 9.37$, $p < .001$, $MSE = .05$. Overall, HRs increased with study time, $F(2, 56) = 80.28$, $p < .001$, $MSE = .03$. There was a condition by study time interaction, $F(6, 168) = 2.93$, $p < .010$, $MSE = .05$, due to the different rate of rise in the HR for the various conditions. HF words studied under the no-task condition showed no numerical improvement between the 500 ms and 1 s condition, replicating prior findings (Estes & Maddox, 2002; Malmberg & Nelson, 2003). A simple ANOVA for the no-task condition, however, did not support the seeming WF by study time interaction in Figure 2B, $F(2, 56) = 1.17$, $p < .316$, $MSE = .05$. Most importantly, there was a WF by condition interaction, $F(3, 84) = 2.88$, $p < .041$, $MSE = .07$. Tests of simple main effects confirm what is apparent in Figure 2B. The LF-HR advantage is present only for the no-task condition, $F(1, 28) = 11.35$, $p < .002$, $MSE = .05$, and disappears for the three study tasks used here, concreteness, vowel counting, and pleasantness (all Fs < 1.33). No other main effects or interactions approached significance (all Fs < 1.75).

Experiment 5

Previous research has identified the importance of correlations of orthographic differences with word frequency. Malmberg and Nelson (2003) proposed that lexical access to LF words requires extra attentional resources, because such words have atypical orthographic structure. According to the REM model (Shiffrin & Steyvers, 1997), feature distinctiveness determines the amount of evidence contributed by a matching feature to the overall evidence that the test item is a target. Such hypotheses are supported by findings that words with uncommon letters are recognized better than words with common letters when word frequency is held constant (Malmberg, Steyvers, Stephens, & Shiffrin, 2002), and words rated by participants as orthographically distinct are better recognized than words rated less distinct (Zechmeister, 1972). These results helped lead us to use a task requiring participants to judge whether a word contained more vowels than consonants (vowel counting) in the hope that this task would focus attention on orthographic features. Yet vowels are generally quite common, and such a judgment does not necessarily require that the identities of the consonants be accessed, because it is possible to make judgments by comparing number of vowels with word length. Thus, we added two new tasks to Experiment 5 that we hoped would explicitly draw attention to unusual letters in a final attempt to find a condition in which the LF-HR advantage would reappear.

Method

A total of 46 Indiana University students participated as part of a course requirement. The design was identical to Experiments 3 and 4 with one exception. The study tasks were (a) concreteness, (b) no task, (c) *spelling* (“Prior to this experiment, did you know the spelling of this word?”), and (d) *unusual letters* (“Does this word contain any unusual letters?”).

Results

Yet again, FARs to HF words were greater than LF words, $t(45) = 7.32$, $p < .001$, $SEM = .01$, as shown in Table 1. Figure 2C gives the HRs. A $2 \times 3 \times 4$ repeated measures ANOVA (word frequency, study time, and condition) showed that HRs increased with study time, $F(2, 90) = 39.36$, $p < .001$, $MSE = .06$. The main effect of WF, $F(1, 45) = 11.21$, $p < .002$, $MSE = .07$, is qualified by a WF $\times$ Task interaction, $F(3, 135) = 2.97$, $p < .034$, $MSE = .04$. Tests of simple main effects confirmed what is apparent in Figure 2C: An LF-HR advantage is present for the no-task condition (replicating Experiments 1 and 4), $F(1, 45) = 4.93$, $p < .031$, $MSE = .05$, and present for the unusual letters task, $F(1, 45) = 22.65$, $p < .001$, $MSE = .04$. There was no difference between the HRs of LF and HF words for the concreteness (replicating all the earlier studies) or spelling tasks (all Fs < 1). Study time and WF produced an interaction, $F(2, 90) = 3.64$, $p < .030$, $MSE = .04$, as did study time and condition, $F(6, 270) = 2.79$, $p < .012$, $MSE = .04$. Both interactions seemed to be due to the different patterns between those conditions that resulted in an LF-HR advantage and those that did not. HRs increased monotonically and similarly for HF and LF for the spelling and concreteness tasks. However, for the unusual letters and no-task conditions, the HRs for LF and HF began together and then separated as study time grew. No other main effects or interactions approached significance (all Fs $<$ 1.69).

General Discussion

These experiments explored in some detail the effects of study task and study time on the WFE generally, and the LF-HR advantage in particular. The first findings concern false alarms: Regardless of the particular task(s) at study and whether they are mixed within a list or not, FARs are consistently 9% to 13% higher for HF than LF words. Note that all of our tasks demonstrate an overall recognition advantage for LF words, even when the HRs do not differ, because all conditions show an LF-FAR advantage. Thus, our arguments and explanations are aimed to explain the changes in size of the LF advantage, not its presence or absence.

Next consider the effects of study time and study task on HRs. Most of our studies used mixed list designs, in which there is a single FAR; thus, performance differences (measured by any method for combining hits and false alarms) are equivalent to HR differences. Not surprisingly, additional study time improves performance for both HF and LF words. Such improvement is fairly large, monotonic, and independent of frequency (HRs were roughly equal for LF and HF words) for the tasks we labeled concreteness, animacy, pleasantness, frequency, spelling, and vowel counting. However, for the no-task and unusual letters conditions, the pattern was different: (a) HRs were higher for LF than HF, averaging over all study times; (b) for brief study times, HRs for HF and LF words tended to be fairly close (although LF words were a bit higher); and (c) HRs for LF words tended to improve more with study time than did HRs for HF words. These generalizations summarize our findings, although we note that a few inconsistencies cropped up between studies, as illustrated in Figures 1 and 2.

The general increase in HRs with study time matches those from many studies. The fact that study time interacts with frequency for the no-task condition also generally matches previous findings, but there are some inconsistencies in the details. For example, Malmberg and Nelson (2003) showed that the magnitude of the LF-HR advantage in no-task conditions does not change after approximately 1 s of study. This result partially matched the results from our Experiment 1 (and the findings from the unusual letter task in Experiment 5), but we did find an increase in the LF-HF–HR difference between 1 and 3 s in the no-task condition of Experiment 5. Experiment 4 lies in the middle with a numerical but not statistically reliable increase in the LF-HF difference from 1 s to 3 s.

Our findings (with one exception) replicate scattered findings showing that certain study tasks produce a reduction or elimination of the LF-HR advantage (Guttentag & Carroll, 1997; Hilford et al., 1997; Hirshman & Arndt, 1997; Mandler et al., 1982). We replicated this outcome for the concreteness, animacy, pleasantness, frequency, spelling, and vowel-counting tasks. An exception occurred for the unusual letter task. This task produced a considerable LF-HR advantage, providing an important constraint for theorizing. Claiming that encoding strategies and encoding processes are an important contributor to the WFE is almost a tautology given that mixing of study tasks within list produces markedly different WFE patterns. To argue that such differences arise at retrieval would require the identification of a test item as having been studied in a particular task in order to apply different retrieval operations. Such identification presupposes that the item is old, and therefore cannot sensibly be used as a starting point for making an old–new decision. It is important to note that an inference that encoding effects are responsible for the differences we observe across tasks does not preclude the possibility that retrieval factors also play an important role in the WFE (cf. Hintzman et al., 1994).

We now turn to various hypotheses, models, and theories that have been proposed to account for the WFE in recognition. The early-phase elevated-attention hypothesis of Malmberg and Nelson (2003) assumes that LF words attract more attention during the early stages of processing, stages during which the lexical entry of the study word is accessed. This extra attention results in more features or more distinctive features being stored for LF words. According to this hypothesis and as evident in the data, the benefit develops over a second or so and then remains stable. The stability of the magnitude of the LF-HR benefit for longer study times presumably implies that after approximately 1 s of study, encoding switches to a form in which LF and HF words benefit equally (possibly semantic or associative). All this reasoning applies to the standard situation in which participants are told to study for a later memory test.

How would Malmberg and Nelson (2003) account for elimination of the LF-HR advantage when participants were asked to carry out a study task? When discussing the findings of Hirshman and Arndt (1997) and Hoshino (1991), they proposed that such tasks cause a selective benefit for HF words, a selective benefit that occurs in the late stage of processing. They could account for the current data in the same manner, but the approach is very descriptive. It is hard to see how a prediction could have been made in advance of the data that the no-task and unusual letter conditions would produce the usual mirror effect, but concreteness, vowel counting, and others would eliminate the LF-HR advantage. More importantly, this hypothesis seems to predict that an LF-HR advantage should be present for the shortest study times and then disappear for longer study times, contrary to the data.

Attention likelihood theory (ALT; Glanzer & Adams, 1990) assumes that LF words receive more attention than HF words, and consequently more LF features are marked at study. The system takes this into account by comparing the expected number of marked features with the actual number of marked features in a likelihood ratio. Both reducing study time and the use of less-effective study tasks are modeled as a reduction in the number of marked features, and thus harm performance by reducing the HR and increasing the FAR (Hilford et al., 1997; Kim & Glanzer, 1993). As noted by Hilford et al. (1997), "Conditions that impair or hinder recognition performance cause a symmetric movement of both the old and the new underlying distributions" (p. 594). The previous conclusions were all based on the standard ALT model in which the system is assumed to estimate the expected number of marked features on the basis of the class of the test item (i.e., there are different expected values for LF and HF words). As pointed out by Stretch and Wixted (1998), ALT can predict HF-FAR $>$ LF-FAR along with an HF-HR advantage under the following conditions: More HF than LF features are marked during study, and the expected number of marked features is assumed to be the average of the test items (i.e., the expected values are equal for HF and LF words). To fit the current set of data, ALT must assume that an equal or greater number of HF features are marked for all tasks except for the unusual letter and no-task conditions. Further, they must assume that participants change their estimate of the expected number of marked features on a trial-by-trial basis depending on the presumed condition of the test item. This seems to require a source judgment to determine the test item's study task, a source judgment that seems to presuppose recognition itself. Such an approach is unsatisfactory not only for this reason, but also because it seems inconsistent with previous articles proposing that orienting tasks produce concentering (i.e., Hilford et al., 1997).

There are many dual-process models that vary in their exact implementation and details (see Yonelinas, 2002, for a review). We limit our discussion to those dual-process models that propose the following general mechanism for the WF mirror effect (Guttentag & Carroll, 1997; Joordens & Hockley, 2000; Reder et al., 2000). Specifically, consider models that assume the FAR portion of the WFE is due to the higher baseline familiarity of HF words and that this initial HF advantage is overcome by the stronger recollection of LF targets. Joordens and Hockley (2000) explained the within-list strength effect by assuming that extra study of HF words equates their encoding and ability to be recollected with LF words. Likewise, they argued that Hirshman and Arndt's (1997) concreteness task equates the participants' ability to recollect HF and LF words. This same explanation can be applied to the current experiments, but it remains a question of why and how these tasks help the ability to recall HF words without benefiting LF words. This explanation is further complicated by the claim that conditions reducing recollection (e.g., an incidental study list with many buffers, Joordens & Hockley, 2000; drug-induced amnesia, Hirshman et al., 2002; speeded testing, Balota et al., 2002) reverse or eliminate the LF-HR advantage. In his review of the literature, Yonelinas (2002) claimed that empirical findings following from the dual-processing framework show that semantic tasks lead to an increase in recollection (and a small increase in familiarity) compared with perceptual tasks. In accordance, the dual-processing framework (as described by Yonelinas, 2002) predicts a decrease in the magnitude of the WFE as the study task becomes more perceptual and less semantic. Our data do not confirm this prediction. We only find the full mirror pattern when participants are looking for unusual letters during study or when performing no task at all.

The source of activation confusion dual-process model (SAC; Reder et al., 2000) differs slightly from those above. SAC assumes that HF words have higher baseline familiarity and thus have a higher FAR. Because LF words have been seen in fewer contexts, any particular event (e.g., the study episode) is more likely to be recollected; thus, there is a higher HR for LF words. However, this model differs from those described above because it predicts a WFE for both conceptual and perceptual aspects (e.g., see Reder, Donavos, & Erickson, 2002). Still, SAC has no a priori basis for predicting the elimination of the LF-HR advantage when words are studied under various study tasks (as demonstrated in the present experiments). Further, SAC predicts a strength-based mirror pattern: As performance for targets improves and the HR increases, FARs decrease, because participants adjust their criterion according to the perceived ease of recollection (Cary & Reder, 2003). This prediction is inconsistent with our data. Across all groups of participants, we found a reliable difference in the overall HRs but no difference in FARs.[2] Thus, although SAC may be more likely to produce a WFE based on perceptual information, it is not yet capable of handling the current set of data.

The REM (Shiffrin & Steyvers, 1997) theory assumes that the strength of a test item is based not only on the fact that it shares features with studied items but also on the diagnosticity (i.e., distinctiveness) of these matching features. A match of an uncommon, more diagnostic feature provides higher evidence that the test item was studied than a match of a common, less diagnostic feature. How does this approach handle word frequency? HF words are assumed to have more common features than LF words. Consider first the implications for FARs. Foils only match other studied words by chance. HF foils therefore tend to have more matches because common features match by chance more often than rare features. Thus, HF-FARs exceed LF-FARs. When a target was tested, one of the traces in memory was stored because of study of that word. The features of this target trace match because they were stored correctly during study. Thus, the evidence for matching for an LF target trace will be based on matching less common and more diagnostic features and will exceed the evidence for HF target traces. Of course all traces other than the target trace will still match only by chance. Thus, the matching evidence for a target test is a combination of evidence from the target trace, favoring LF words, and the matching evidence from the many other traces, favoring HF words. In normal situations, this balance is resolved in favor of better evidence for LF words, producing an HR advantage for LF words. The decision rule is based on likelihood ratios, which naturally produce a mirror effect and concentering.

There are several ways within this framework that one might model the effects of different study tasks and study times. Perhaps the simplest idea is to assume that the specified study tasks move attention away from the uncommon features of LF words during early processing stages when uncommon (visual) features are encoded, thereby reducing the LF-HR advantage. According to this view, the use of the task requiring judgment of unusual letters places attention back on the rare features and does so especially late in processing, thereby reintroducing the LF-HR advantage. To explain why the vowel counting task does not produce the same effect, one would have to argue that this task focuses attention on vowels (in contrast to consonants) and that vowels tend to be common.

The opposite approach is also possible, in which the specified tasks selectively cause extra processing of HF words, thereby raising their HRs and eliminating their normal HR disadvantage. Such an approach makes sense for our semantic study tasks (although leaving open the question as to why HF words do not benefit similarly from instructions to study for a later memory test) but does not provide a compelling explanation for the reappearance of the LF-HR advantage in the unusual letters task.

In summary, to account for the elimination of the LF-HR advantage, most of the above approaches assumed that certain study tasks disproportionately benefited the encoding of HF words in comparison with normal instructions to study for a later memory test. Such a view is consistent with results showing that participants free to allocate time spent studying words of varying frequency spend more time on LF than HF words (Rao & Proctor, 1984). Perhaps this is due to their misconception that LF words will be harder to remember (Wixted, 1992). The majority of tasks that we used and the limited study time may eliminate this selective attention (assuming processing ends when the task is complete), thereby harming LF encoding or helping HF encoding, relative to normal study. According to this approach, the unusual letter task is an exception to the other tasks because it focuses attention selectively on LF words.

Let us turn to a more general look at the effects of study tasks on encoding. When participants study lists under normal recognition instructions, they attend primarily to visual information and surface information (e.g., Mandler, 1980; Schulman, 1967). Providing a task could induce a shift from this default strategy to a deeper type processing as required by the task. This idea is indirectly supported by studies showing that standard instructions and surface-level study tasks lead to about equal recognition performance (Elias & Perfetti, 1973). Similarly, incidental study harms recall but not recognition (Eagle & Leiter, 1964). Likewise, when studying for an item recognition task, performance on an unexpected associative recognition task is harmed compared with performance when expecting an associative recognition task (Hockley & Cristi, 1996). These findings indicate that participants do not routinely engage in extensive controlled processing during a typical experimental study list. On the basis of this, one might expect an interaction of the provided study task with study time (as we observed in Experiments 1, 4, and 5). Once the required task is completed, which in some cases could be well ahead of the study time provided, the participant would have little reason to continue processing of any sort. Although this idea has merit, it needs more elaboration to explain adequately the present results.

In conclusion, the present results do point to an important role for feature distinctiveness in the WFE. As noted earlier, LF words have more visually distinct features than HF words (Kinsbourne & Evans, 1970; Malmberg et al., 2002; Zechmeister, 1969, 1972). The data seem consistent with the view that

[2] Collapsing over study time and study task, the overall HR varied across participant group for LF-HR, $F(5, 199) = 8.60$, and HF-HR, $F(5, 199) = 5.21$, whereas FARs did not vary across groups for neither LF-FAR, $F(5, 199) = 1.46$, nor HF-FAR, $F(5, 199) < 1$.

relatively automatic initial encoding is directed toward visual and surface features that produce a distinctiveness advantage for LF words in all conditions. Standard instructions and instructions to attend to unusual features lead to more storage of such features, whereas instructions to perform other tasks lead to less storage of such features, particularly in the later stages of processing. This hypothesis could be formalized in various ways and inserted into many of the theories described above. Our focus on orthography is not to say that no other factors contribute to the WFE. It is likely that factors such as the number of prior contexts (e.g., Dennis & Humphreys, 2001; cf. Criss & Shiffrin, in press; Reder et al., 2000; Sikström, 2000; Steyvers & Malmberg, 2003) contribute to the WFE.

The fact that most of the study tasks we used eliminated the LF-HR advantage, even when the tasks were mixed within list, points to the importance of encoding processes in the WFE. Hintzman et al. (1994) said "It [the mirror effect] may best be attributed to the inherent nature of the retrieval and judgment processes that underlie recognition memory" (p. 286). The present results qualify this conclusion by pointing to a similarly important role for encoding.

References

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition memory. *Journal of Memory and Language, 46,* 199–226.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Memory & Cognition, 49,* 231–248.

Criss, A., & Shiffrin, R. M. (in press). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*.

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

Dorfman, D., & Glanzer, M. (1988). List composition effects in lexical decision and recognition memory. *Journal of Memory and Language, 27,* 633–648.

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology, 68,* 58–63.

Elias, C., & Perfetti, C. (1973). Encoding task and recognition memory: The importance of semantic encoding. *Journal of Experimental Psychology, 99,* 151–156.

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

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 19,* 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, and Cognition, 16,* 5–16.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567.

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

Hilford, A., Glanzer, M., & Kim, K. (1997). Encoding, repetition, and the mirror effect in recognition memory: Symmetry in motion. *Memory & Cognition, 25,* 593–605.

Hintzman, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 201–205.

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 275–289.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases 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., & Palij, M. (1992). Rehearsal and the word frequency effect in recognition memory. *Journal of Memory and Language, 31,* 477–484.

Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition, 24,* 202–216.

Hoshino, Y. (1991). A bias in favor of the positive response to high-frequency words in recognition memory. *Memory & Cognition, 19,* 607–616.

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, and Cognition, 26,* 1534–1555.

Kim, K., & Glanzer, M. (1993). Speed versus accuracy instructions, study time, and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 638–652.

Kinsbourne, M., & Evans, N. (1970). Is the word frequency effect on recognition threshold a function of transitional probability? *Psychological Reports, 27,* 143–145.

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

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,* 630–644.

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. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

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

McClelland, J. L., & Chappell, M. (1999). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 734–760.

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

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

Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition, 30,* 312–323.

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.

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

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

Sikström, S. (2001). The variance theory of the mirror effect in recognition memory. *Psychonomic Bulletin & Review, 8,* 408–438.

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

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24,* 1379–1396.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 681–690.

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. (1969). Orthographic distinctiveness. *Journal of Verbal Learning and Verbal Behavior, 8,* 754–761.

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

Received July 24, 2003
Revision received November 20, 2003
Accepted December 27, 2003 ■