Memory & Cognition
2006, 34 (7), 1495-1511

Perception and recognition memory of words and werds: Two-way mirror effects

D. VAUGHN BECKER, STEPHEN D. GOLDINGER, and GREGORY O. STONE
Arizona State University, Tempe, Arizona

We examined associative priming of words (e.g., TOAD) and pseudohomophones of those words (e.g., TODE) in lexical decision. In addition to word frequency effects, reliable base-word frequency effects were observed for pseudohomophones: Those based on high-frequency words elicited faster and more accurate correct rejections. Associative priming had disparate effects on high- and low-frequency items. Whereas priming improved performance to high-frequency pseudohomophones, it impaired performance to low-frequency pseudohomophones. The results suggested a resonance process, wherein phonologic identity and semantic priming combine to undermine the veridical perception of infrequent items. We tested this hypothesis in another experiment by administering a surprise recognition memory test after lexical decision. When asked to identify words that were spelled correctly during lexical decision, the participants often misremembered pseudohomophones as correctly spelled items. Patterns of false memory, however, were jointly affected by base-word frequencies and their original responses during lexical decision. Taken together, the results are consistent with resonance accounts of word recognition, wherein bottom-up and top-down information sources coalesce into correct, and sometimes illusory, perception. The results are also consistent with a recent lexical decision model, REM-LD, that emphasizes memory retrieval and top-down matching processes in lexical decision.

Studies of word perception are often framed in terms of signal detection theory (SDT; Green & Swets, 1966), especially studies focused on lexical decision (Balota & Chumbley, 1984) or semantic priming (Rhodes, Parkin, & Tremewan, 1993). SDT provides a metaphorical description of decision making (including lexical decisions) and methods for data analysis. Moreover, SDT provides a conceptual framework that separates the "bottom-up" collection of sensory information from "top-down" decision processes that follow. In the word perception literature, most studies focus on response time (RT) data, using such tasks as speeded naming or lexical decision. RT is typically the primary measure in lexical decision because accuracy is high, although modified methods focus on error rates (Hintzman & Curran, 1997). Nevertheless, in experimental procedures involving two-alternative classification, as in lexical decision, SDT can help estimate sensitivity and bias, potentially guiding the interpretation of some experimental manipulation. And even when performance is highly accurate, the SDT framework may help explain RT patterns (Balota & Chumbley, 1984; Gordon, 1983; Lewellen, Goldinger, Pisoni, & Greene, 1993).

For example, Schvaneveldt and McDonald (1981) examined semantic priming across three detection tasks, examining both errors and RTs. Different tests were chosen, focusing participants' attention to different "levels" of lexical analysis: Some participants tried to detect gaps in single letters in words, others looked for rotated letters within words, and others made lexical decisions. In every task, words and nonwords were preceded by neutral, unrelated, or related primes. Also, each task involved either tachistoscopic (50-msec) stimulus presentation or longer, response-terminated presentation. The most interesting results arose in lexical decision: Participants discriminated words (TIGER) from nonwords that differed by single, noninitial letters (TIGAR). Semantic priming sped up lexical decisions when targets remained in sight. With brief exposures, however, priming mainly increased false alarms (FAs) to nonwords. Schvaneveldt and McDonald suggested that a *verification process* occurs in lexical decision (cf. Paap, Newsome, McDonald, & Schvaneveldt, 1982). Specifically, they proposed that an early stage of word perception is purely bottom-up, as visual features are extracted. Top-down processes, characterized as spelling verification, are relatively late arriving (as in the cascade model; McClelland, 1979).

Lexical Decision As Signal Detection

When characterizing the lexical decision task in SDT, it is commonplace to describe the words and nonwords as "signals" and "noise," respectively. Collectively, items give rise to distributions of psychological evidence: Their degree of overlap affects error rates and may also affect RTs (Balota & Chumbley, 1984). Thus, nonword charac-

This work was supported by NIDCD Grant R01-DC04535 to S.D.G. We thank Guy Van Orden for many useful discussions. We also thank Michael Masson and two anonymous reviewers for suggested improvements. Correspondence may be addressed to any of the authors, at Department of Psychology, Box 871104, Arizona State University, Tempe, AZ 85287-1104 (e-mail: vaughn.becker@asu.edu, goldinger@asu.edu, or greg.stone@asu.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

teristics strongly affect lexical decision performance. Discriminating NURSE from XLFRT is easy, requiring superficial assessment of orthographic or phonologic structure. Pronounceable nonwords such as GERSE make the task harder, requiring true phonologic (and perhaps semantic) discrimination. Pseudohomophones such as NERSE are the most challenging: With meaningful pronunciations, these foils must be rejected solely on semantic–orthographic mismatch. Each increase in word–nonword similarity increases the signal–noise overlap, reducing accuracy and slowing RTs (see Figure 1). When nonword foils are held constant, manipulations of the signal distribution, such as varying mean word frequencies, can create the same pattern. Stone and Van Orden (1993) found both patterns: Increasing the difficulty of lexical discrimination slowed overall responding and dramatically increased word frequency effects.

In addition to changing sensitivity, various manipulations, such as altering the proportions of words and nonwords, can affect decision criteria. If 80% of items are pronounceable nonwords, low-frequency words such as ASP will likely be rejected (missed). This criterion would reflect overall list composition, but criterion shifts can also be induced on individual trials, perhaps by associative priming (Neely, 1991). Seeing DOCTOR before NURSE may bias participants to (correctly) respond "word." To fully address lexical decision, some theories postulate two decision criteria, reflecting different memory processes. Balota and Chumbley (1984; see Atkinson & Juola, 1973) proposed that psychological distributions for words and nonwords arise along a *familiarity/meaningfulness* (F/M) dimension, with two criteria demarcating a central region of overlap (as in Figure 1). Nonwords with F/M values falling below the lower criterion are quickly rejected. Words with F/M values exceeding the upper criterion are quickly accepted. Stimuli with F/M evidence in the central region require extra analysis, increasing decision times. This region is mainly populated by uncommon words and "word-like" nonwords.

Given this framework, it is natural to expect low-frequency words (e.g., SNAIL) to create weak F/M signals, relative to high-frequency words (e.g., CHAIR). However, the proper expectation for pseudohomophones is unclear. Which nonwords (SNALE or CHARE) should fall closer to the central region? This question is challenging because few word-perception models or theories directly address nonword rejection. Instead, they typically model the accumulation of evidence for "word" responses (i.e., the traversal from letter string presentation to unique lexical access), with well-defined processes that predict effects of word frequency, consistency, priming, orthographic neighborhoods, and other variables (e.g., Seidenberg &

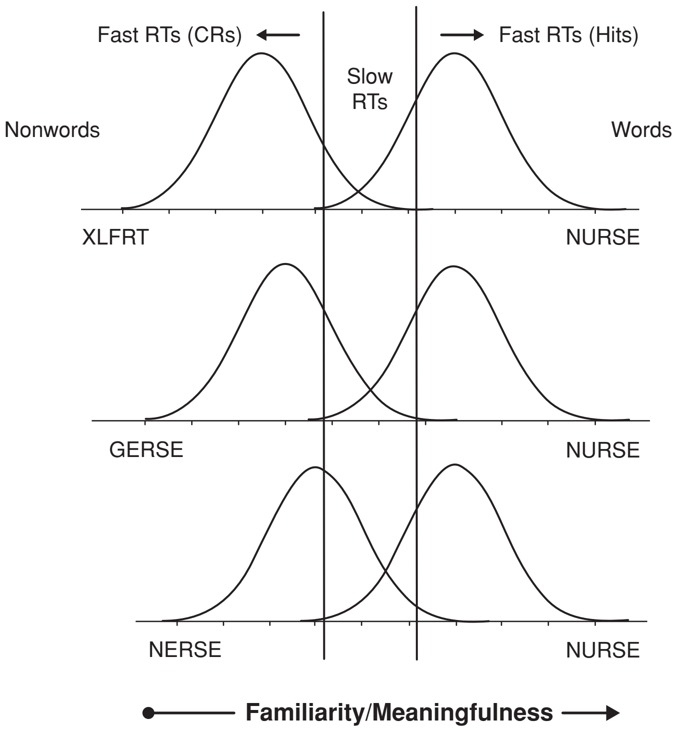

Figure 1. Hypothetical psychological distributions for nonwords and words, arrayed along a familiarity/meaningfulness dimension. Overlap between the distributions increases (reducing sensitivity) as nonwords become more word-like. Two decision criteria are hypothesized, helping explain common response latency results (Balota & Chumbley, 1984; see text).

McClelland, 1989). By contrast, "nonword" responses are modeled passively, as default responses given failures of timely lexical access.

Consider two well-known models of lexical access, the multiple readout model (MROM; Grainger & Jacobs, 1996) and the dual-route cascaded (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), both of which operate according to similar principles. In both models, word perception occurs by gradual activation of sublexical, then lexical, units. Positive "word" responses are elicited if any specific word surpasses a recognition threshold. These word-specific thresholds are relatively fixed, determined by a lifetime of experience, the nature of competitors, and so on. Alternatively, because unique lexical access is not always required, "word" responses may occur when summed activation across many lexical units surpasses a decision criterion. This criterion is context sensitive, subject to strategic control in lexical decision, and can thus explain changes in "word" RTs based on changes in nonword properties. Finally, "nonword" responses are generated with reference to a third criterion, a temporal deadline *T* (see Wagenmakers, Steyvers, et al., 2004). If neither positive threshold is surpassed by time *T*, the participant responds "nonword." This threshold is also subject to strategic control, also potentially explaining nonword property effects and speed–accuracy trade-offs.

In terms of global cognitive architecture, both the MROM and the DRC model generate lexical decisions entirely by "bottom-up" processes: They discriminate words and nonwords without any explicit decision stage. From this bottom-up perspective, the pseudohomophones SNALE and CHARE are equally unfamiliar letter strings, but CHARE has a more familiar phonological pattern. Phonological activation is a powerful source of lexical evidence in both models. Thus, CHARE will elicit a strong F/M signal, leading to slow rejections and occasional FAs, relative to the less familiar pattern SNALE. This analysis was verified by Ziegler, Jacobs, and Klüppel (2001), who conducted simulation tests of both MROM and DRC: Both models predicted slower nonword rejections for pseudohomophones based on higher frequency words; performance to SNALE was predicted to exceed performance to CHARE. However, Ziegler et al. (2001) noted a prior study (Van Orden et al., 1992) that showed a small (16-msec) effect in the opposite direction. Ziegler et al. (2001) replicated the effect in German, extending it across different word lengths.

Having verified that base-word frequency effects contradict both the MROM and DRC models, Ziegler et al. (2001, p. 552) posed the question: "So what is wrong with the models?" They argued that, when lexical decisions are modeled using only bottom-up activation levels, slower rejections will always be predicted for more "familiar" nonword strings. Conversely, from a top-down perspective, as in spelling verification, high-frequency pseudohomophones such as CHARE should present a salient mismatch between semantics and orthography. Thus, CHARE would elicit a weak F/M signal and should be easily rejected, relative to low-frequency pseudohomophones such as SNALE. In the present research, we contrasted low- and high-frequency pseudohomophones, in a manner similar to Van Orden et al. (1992) and Ziegler et al. (2001).

Associative Priming and Phonologic Coherence

One goal of the present investigation was to test the *phonologic coherence hypothesis*, derived from a general resonance framework for word perception (Van Orden & Goldinger, 1994; see also Gottlob, Goldinger, Stone, & Van Orden, 1999). According to this hypothesis, orthographic–phonologic (O–P) dynamics stabilize earliest in visual word recognition, providing a coherent basis to stabilize ongoing, higher order semantic (S) processes. Phonology plays such a central role because of *covariant learning*. Spellings and sounds covary almost perfectly in some languages (e.g., Serbo-Croatian). In English, most consonant graphemes (e.g., *d*, *m*, *g*) denote one or two phonemes across words; vowel graphemes (e.g., *e*, *o*, *u*) denote four to six phonemes across words. Given such tight covariation, models that learn statistical mappings easily "discover" O–P relations. By contrast, mappings between spellings (or sounds) and semantics are far less consistent. For example, a word-initial *d* is perfectly correlated with the phoneme /d/, but both surface forms are shared across thousands of lexical entries (*dog*, *diet*, *druid*, etc.). Although a *d* clearly denotes the phoneme /d/, it provides almost no information about the intended meaning (e.g., *dog*).

Van Orden and Goldinger (1994) discussed the phonologic coherence hypothesis by reference to a general *resonance* framework, consistent with many theories of perception and memory. In network models, resonance is achieved when feed-forward and feedback sources of activation are mutually reinforcing (Grossberg & Stone, 1986). Indeed, the principle of phonologic coherence may naturally emerge from many connectionist models of reading, such as the well-known "triangle" model of Seidenberg and McClelland (1989; see Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Rueckl, 2002; Seidenberg & Zevin, 2003). In word perception, presentation of a letter string is assumed to send a diffuse wave of activation from orthographic "nodes" to all associated phonologic and semantic "nodes." Once activated, these higher order nodes return feedback to the orthographic nodes (Stone, Vanhoy, & Van Orden, 1997). If the feedback pattern is a reasonable match to the initial stimulus pattern (degrees of activation between nodes are commensurate with their previous covariation), the cycle is repeated. Within limits, this resonance is self-perpetuating, binding the separate knowledge sources into a coherent perceptual experience (Grossberg, 1980).

In the present research, we compared lexical decision performance to low- and high-frequency words. The words were preceded by either associated or unrelated primes. Of critical importance, each word had a pseudohomophone foil, which was either primed or not, depending on the experiment. Prior research (e.g., Lukatela, Eaton, Lee, Carello, & Turvey, 2002; Lukatela & Turvey, 1994a, 1994b) has shown that interword priming is strongly affected by phonologic coherence. For example, pseudohomophones prove very effective as associative primes (e.g., NERSE–DOCTOR), working as effectively as their base words. Even identity priming (priming a word by itself) is reduced by high O–P ambiguity (Lukatela, Frost, & Turvey, 1999), suggesting that the coherence of O–P mappings plays a dominant, early role in word perception.

With this principle in mind, the present experiments were intended to explore the processes that follow early O–P coherence. Following Van Orden et al. (1992; Ziegler et al., 2001), we examined performance in a context wherein fast O–P resonance could not predict true lexical status, so nonword rejection must occur via the O–S resonance. We extended prior research, however, by including associative primes: For real words, especially low-frequency words, priming should improve performance (Becker & Killion, 1977; Neely, 1991). The potential priming effects to pseudohomophones were of greater interest: According to bottom-up theories (e.g., MROM or DRC), primed pseudohomophones should be difficult to reject, relative to an unprimed condition. According to a top-down approach, the expectation is rather unclear. For example, the bottom-up prediction could be substantiated—even top-down theories begin with bottom-up signals; stronger signals may indeed be more difficult to reject. Conversely, by boosting semantic activation, priming may increase the efficiency of O–S resonance or verification, leading to improved nonword rejection. (Note that neither bottom-up theory can make this prediction.) In essence, the proper prediction is a matter of tipping scales: Which direction of processing will be more affected by priming? This question, in turn, is further complicated by pseudohomophone frequency. By the bottom-up theories, all pseudohomophones should be harder to reject after priming, with more profound priming effects for lower frequency pseudohomophones (following many studies; e.g., Neely, 1991).

In contrast, by the top-down theories, we cannot easily predict the direction of a potential priming effect. Indeed, we might predict divergent effects, based on pseudohomophone frequency. Assume that priming increases the bottom-up F/M signal strength for all nonwords, inhibiting fluent rejection. As MROM and DRC predict, this penalty should be stronger for low-frequency pseudohomophones. By the top-down view, however, this penalty will be partially offset by improved spelling verification (or O–S resonance). Prediction now becomes a matter of degree: For high-frequency pseudohomophones, small bottom-up priming penalties may be offset by large top-down priming benefits, leading to faster, more accurate rejections. For low-frequency pseudohomophones, larger bottom-up penalties may be harder for top-down benefits to offset. In fact, in the absence of explicit model simulations, we could reasonably predict positive, negative, or null priming effects for low-frequency pseudohomophones. In Experiments 1A, 1B, 2A, and 2B, we collected these critical data. In addition to standard RT analyses, we conducted SDT analyses, testing whether any potential differences across high- and low-frequency pseudohomophones were better explained as sensitivity or bias effects.

In addition to lexical status, frequency, and priming, we manipulated list structures in Experiments 1A, 1B, 2A, and 2B. In lexical decision, a "frequency blocking" effect is often observed, wherein word frequency effects change as a result of list composition. Glanzer and Ehrenreich (1979; see also Gordon, 1983) compared lexical decision times in conditions with mixed-frequency lists and in lists with frequency blocking (i.e., all low frequency or all high frequency). Using legal nonwords (e.g., GERSE) as the foils, these authors found that high-frequency words benefited from blocked presentation, but low-frequency words were unaffected. In a later study, Stone and Van Orden (1993) tested frequency blocking using both legal nonwords and pseudohomophone foils. With legal nonwords, they replicated the Glanzer and Ehrenreich pattern. With pseudohomophones, however, the pattern reversed: Low-frequency words showed far larger benefits from frequency blocking. This pattern is difficult to explain, even in terms of shifting response criteria (Stone & Van Orden, 1993). Although we did not use legal nonwords in the present experiments, we compared mixed- and blocked-frequency conditions, testing whether such asymmetric benefits would also arise among low- and high-frequency pseudohomophones.

EXPERIMENTS 1A, 1B, 2A, AND 2B

We first conducted four lexical decision experiments, which are described together in the interest of brevity and clarity. All four experiments had 2 × 2 designs, contrasting frequency and priming manipulations. All experiments also shared basic procedures: On every trial, primes (either associated or unrelated) were shown for 1 sec, followed by response-terminated targets. The only differences concerned priming relations and list structures: In Experiments 1A and 1B, half the word targets were preceded by associated primes; pseudohomophones were always preceded by unrelated primes. In Experiment 1A, frequency was blocked, such that each participant saw only low- or high-frequency items. In Experiment 1B, the participants saw target lists of mixed frequency. Experiments 2A and 2B differed only by the addition of associative priming of the pseudohomophones (e.g., DOCTOR before NERSE), as well as the words.

Method

Participants. A total of 192 Arizona State University undergraduates participated for course credit. All were native English speakers with normal or corrected-to-normal vision. The participants with errors in excess of 33%, or whose mean RTs were 2.5 standard deviations above their group means, were excluded from analyses. Altogether, 13 participants were excluded, leaving 61 participants in Experiment 1A, 28 in Experiment 1B, 58 in Experiment 2A, and 32 in Experiment 2B. Note that Experiments 1A and 2A had larger samples because word (and pseudohomophone) frequencies were manipulated between groups.

Stimulus materials. Monosyllabic words of high and low frequency (≥90 per million and ≤10 per million, respectively; Kučera & Francis, 1967) were selected, following several constraints. Most important, each word had a potential misspelling that was a perfect phonologic impostor. Feed-forward consistent pseudohomophones were first generated using a database of spelling–sound correspondences (Ziegler, Stone, & Jacobs, 1997). Their pronunciations were verified by 20 students; only items with 100% agreement were retained. To ensure that all base words had familiar spellings, another 20 students were given the pseudohomophones in a long list of real words and were asked to identify misspelled items. Only pseudohomophones with ≥90% correct rejections were retained for experimental use. There were 128 items in the final set of pseudohomophones, with 64 based on low- and high-frequency words, respectively (see Appendix A).

Once the pseudohomophones were selected, associative primes for all base words were generated by the first author, mainly using published lists. Prime–target pairs were rated for relatedness: On a 1–7 scale, 40 students indicated how likely each prime would make them think of its target (7 = *very likely*). Unrelated word pairs were included to anchor the ratings. Only word pairs with mean ratings ≥4 were used in the experiments. The final two sets of pseudohomophones (based on low- and high-frequency words) had equivalent orthographic similarity to their base words, according to Van Orden's (1987) measure ($M_{\text{low}} = .691$, $SD = .115$; $M_{\text{high}} = .681$, $SD = .127$). During the experiments, base words and their pseudohomophones were seen equally often across participants; no participant saw both a base word and its pseudohomophone. In similar fashion, associated and unassociated primes were used equally, allowing each item to serve as its own priming control across participants. In data analyses, we collapsed across both forms of counterbalancing.

Apparatus. The experiments were conducted using PCs in sound-attenuated booths. The display was generated using the DMASTR program, version 2.61 (Forster & Forster, 1996). The participants sat approximately 50 cm away from monochrome monitors. The display used the standard 80 × 25 character set, with stimuli presented in the center of the screen and each letter subtending an approximate visual angle of 22′ horizontal and 50′ vertical.

Design. To ensure that each word and pseudohomophone was seen by an equal number of participants in each priming and blocking context, we created 32 counterbalanced lists, 8 per condition. In Experiments 1A and 1B, words were preceded by associated and unrelated primes (50% each); all pseudohomophones were preceded by unrelated primes. In Experiments 2A and 2B, both words and pseudohomophones were preceded by associated and unrelated primes (50% each). Thus, Experiments 1A and 1B contained a total of 25% related priming trials; Experiments 2A and 2B contained a total of 50% related priming trials.

Procedure. The procedure was explained and demonstrated to each participant, then 20 practice trials were given, followed by a pause for questions. In the experiment proper, each randomized trial began with a central fixation sign (+) for 500 msec. This was followed by a prime word, in lowercase letters, which remained visible for 1,000 msec. The prime was immediately followed by the target, in uppercase letters, which remained until a response occurred or 4,000 msec elapsed. The participants indicated whether the target was a word, pressing the right and left "shift" keys to indicate "word" and "nonword," respectively. RTs were recorded from target onset. A 1,000-msec pause separated trials.

Only negative feedback was given: Following any error, the word *wrong* appeared at the fixation point for 1,000 msec, prior to the pause and the next trial. At the end of the experiment, the participants were debriefed. In all experiments, procedures of data collection and analysis complied with APA ethical guidelines for research.

Results

From the selected set of 128 base words, 10 were removed from analyses because of undetected feed-forward inconsistencies in their pseudohomophones. Another 3 were removed for excessive error rates, leaving 56 low-frequency and 59 high-frequency items. Note that the remaining items still had equivalent orthographic similarity to their base words, according to Van Orden's (1987) measure ($M_{\text{low}} = .689$, $SD = .114$; $M_{\text{high}} = .682$, $SD = .131$); in similar fashion, no other matching variables (e.g., mean prime–target relatedness) were adversely affected by these removals. RTs <150 msec and >2,500 msec (less than 0.5% of all trials) were excluded from analysis. Afterward, mean correct RTs and error rates were calculated for participants and items in each condition. Priming × frequency ANOVAs were conducted on these measures for words (and foils in Experiments 2A and 2B). Priming was always a within-items and within-participants contrast. Frequency was always between items but was between participants in the blocked-frequency experiments (Experiments 1A and 2A) and within participants in the mixed-frequency experiments (Experiments 1B and 2B). Full ANOVA results are listed in Appendixes B and C; only key contrasts are discussed in this section. Tables 1 and 2 display the mean RTs and error rates of all four experiments, along with priming benefits or costs.

Words. Results for the words were broadly consistent with the prior literature (e.g., Becker & Killion, 1977; Neely, 1991). Significant priming benefits for words were seen in latency and accuracy across all experiments, with the sole exception of latencies to high-frequency words in Experiment 1B. Robust word frequency effects were also seen across experiments, although priming benefits for low-frequency words were larger in mixed-frequency blocks. In Experiments 2A and 2B (when both words and pseudohomophones were primed), correct "word" responses were slower (794 msec) than those in Experiments 1A and 1B (735 msec) [$t_2(113) = 3.96$ $p < .01$]. Priming pseudohomophones also elicited an increase in misses, from 5.6% to 6.2% [$t_2(113) = 2.8$, $p < .05$]. Both effects suggest that the participants' response criteria were

Table 1
Mean Response Times (in Milliseconds) and Error Rates (%) to Words and Pseudohomophones in Experiments 1A and 1B

	RT	Error	RT	Error
Blocked Frequency (Experiment 1A)				
	Low-Frequency Group (n = 31)		High-Frequency Group (n = 30)	
Words				
Unassociated primes	832	9.9	714	2.2
Associated primes	746	5.6	662	0.2
Priming	86*	4.3*	52*	2.0*
Pseudohomophones				
Unassociated primes	931	12.6	851	4.4
Mixed Frequency (Experiment 1B, n = 27)				
	Low-Frequency Group		High-Frequency Group	
Words				
Unassociated primes	856	15.4	668	5.5
Associated primes	744	3.9	653	1.6
Priming	112*	11.5*	15	3.9*
Pseudohomophones				
Unassociated primes	887	16.7	844	8.8

*$p < .05$.

Table 2
Mean Response Times (in Milliseconds) and Error Rates (%) to Words and Pseudohomophones in Experiments 2A and 2B

	RT	Error	RT	Error
Blocked Frequency (Experiment 2A)				
	Low-Frequency Group (n = 28)		High-Frequency Group (n = 30)	
Words				
Unassociated primes	928	11.4	751	2.9
Associated primes	841	4.1	685	0.9
Priming effects	87*	7.3*	66*	2.0*
Pseudohomophones				
Unassociated primes	1,015	12.1	867	5.8
Associated primes	1,039	18.2	817	4.3
Priming effects	−24	−6.1*	50*	1.5
Mixed Frequency (Experiment 2B, n = 31)				
	Low-Frequency Group		High-Frequency Group	
Words				
Unassociated primes	911	18.7	718	3.6
Associated primes	798	7.4	676	0.6
Priming effects	113*	11.3*	42*	3.0
Pseudohomophones				
Unassociated primes	890	14.6	861	6.3
Associated primes	925	24.0	844	6.7
Priming effects	−35	−9.4*	17	−0.4

*$p < .05$.

stricter when prime–target associations were no longer diagnostic of lexical status.

Pseudohomophones. In Experiments 1A and 1B, error rates were significantly higher for pseudohomophones based on low-frequency words than for those based on high-frequency words, replicating Van Orden et al. (1992). Low-frequency pseudohomophones also elicited significantly slower correct rejections, a stronger effect than Van Orden et al. observed but comparable to those reported by Ziegler et al. (2001).

Priming pseudohomophones had opposite effects, depending on base-word frequency. Given a pure block of high-frequency base words (Experiment 2A), priming sped correct rejections without increasing FAs (a similar, unreliable trend occurred in Experiment 2B). For low-frequency base words (in both pure or mixed lists), priming tended to slow correct rejections (by 27 msec; p = .07) and significantly increased FAs. Overall, associative priming improved performance to high-frequency pseudohomophones but impaired performance to low-frequency pseudohomophones.

Signal-detection analyses. Experiment 2B also revealed an interesting relationship between misses and FAs to low-frequency items. When preceded by unrelated primes, misses to low-frequency words were higher (18.7%) than FAs to low-frequency pseudohomophones [14.6%; $t_1(30) = -1.99, p < .05$]. However, related priming reduced misses by 11.3% and increased FAs by 9.4% [$t_1(30) = 5.80, p < .01$]. Thus, with low-frequency items, priming seemingly induced a bias toward "word" decisions. To formally assess any changes in sensitivity and bias, we conducted signal detection analyses on lexical decision accuracy. The analyses were restricted to the frequency-blocked conditions because their criterion values would have been unaffected by widely varying frequencies. Both parametric and nonparametric measures were calculated for each participant, separately for related and unrelated priming trials. Differences between these measures were then entered into matched-sample t tests.

The participants occasionally performed perfectly in at least one subcondition. To allow calculation of the indexes, hit rates equal to 1 were lowered to .9375 and FA rates of 0 were raised to .0625 (these values correspond to one half of an error). The parametric sensitivity measure was d'. The parametric bias measure was C, which ranges from -1 to $+1$, with positive values indicating a "word" bias (Macmillan & Creelman, 1991). The nonparametric sensitivity measure was A'; this measure ranges from .5 to 1.0, with 1.0 indicating perfect sensitivity. Finally, the nonparametric criterion measure was B''_D, which is independent of A' (Donaldson, 1992). It also ranges from -1 to $+1$, with negative values indicating a "word" bias. Means for all these measures are shown in Table 3.

According to d', priming increased sensitivity between high-frequency words and their pseudohomophones [$t(29) = 2.10, p < .05$], although this effect was only marginal when tested using A' [$t(29) = 1.62, p = .12$]. In the low-frequency block, priming had no effect on sensitivity by either measure [both ts(29) < 1], although trends emerged toward decreased sensitivity. Despite these trends, neither potential priming × frequency interaction was reliable.

According to C, priming biased the participants to respond "word" in the low-frequency block [$t(25) = 4.13, p < .001$], a finding that was corroborated by B''_D [$t(25) = 4.45, p < .001$]. In the high-frequency block, priming had no impact on criteria, according to either measure [both ts(25) < 1]. This pattern was verified by reliable priming × frequency interactions, using both criterion measures [C, $F(1,54) = 19.3$; B''_D, $F(1,54) = 20.4$].

Discussion

The results of Experiments 1A, 1B, 2A, and 2B were relatively straightforward. Response patterns to words (frequency effects, priming effects, and their interaction) were consistent with many previous studies. We also observed global changes in performance when comparing results from Experiments 1A and 1B with results from Experiments 2A and 2B: When associative priming was no longer diagnostic of lexical status, performance in "word" trials was slower and frequency effects were increased. In short, the "word" data followed common patterns, consistent with many theoretical frameworks.

Of greater interest, the response patterns to pseudohomophones were consistent with the schematic framework shown in Figure 2: The nonword distribution is best depicted as a mirror image of the word distribution, such that high-frequency words are advantaged in acceptance and high-frequency pseudohomophones are advantaged in rejection. In recognition memory, a well-documented pattern is the *mirror effect* (see Glanzer, Adams, Iverson, & Kim, 1993). There are robust word frequency effects

Table 3
Signal Detection Measures From Experiment 2A

Frequency Block	d'	A'	C	B''_D
Low Frequency				
Unassociated	2.43 (.11)	.927 (.01)	−.008 (.04)	−.183 (.07)
Associated	2.36 (.12)	.917 (.01)	.248 (.05)	.362 (.07)
High Frequency				
Unassociated	2.86 (.07)	.954 (.01)	.058 (.03)	.091 (.04)
Associated	2.98 (.04)	.961 (.01)	.043 (.02)	.069 (.03)

Note—Standard errors are shown in parentheses.

in recognition memory, but they are directly opposite to the word frequency effects in perception: Low-frequency words typically lead to better recognition memory, perhaps reflecting their distinctive nature, relative to more common words (Shiffrin & Steyvers, 1997). The mirror effect involves low-frequency words conveying a *double bonus*, yielding more hits and fewer FAs. In Experiments 1A, 1B, 2A, and 2B, we consistently observed a perceptual version of the mirror effect, with a similar double bonus for high-frequency items. As noted earlier (and by Ziegler et al., 2001), this pattern cannot naturally emerge from bottom-up theories such as MROM or DRC. In these theories, the prediction for low- and high-frequency pseudohomophones is clear: Higher frequency pseudohomophones resemble more common words, sound identical to more common words, and are occasionally primed with associates of more common words. Whether by thought exercise or simulation, MROM and DRC predict that higher frequency pseudohomophones will generate a stronger "word" signal along the F/M dimension (Balota & Chumbley, 1984).

Despite this prediction, the opposite result consistently emerged: The participants were better at rejecting higher frequency pseudohomophones, even when tempted by priming to respond "word." Similar patterns have been noted in studies of semantic classification tasks, since lower frequency homophones or pseudohomophones are more difficult to reject (e.g., correctly rejecting ROWS or ROZE as a member of the category FLOWER; see Jared & Seidenberg, 1991, and Van Orden, 1987). As noted by Ziegler et al. (2001), this pattern is consistent with theories that entail top-down decision processes, such as a spelling verification stage (Paap et al., 1982). In our view, the results are consistent with two theories. As we later note, these are not competing views but are theories that naturally complement one another.

First, as discussed earlier, the lexical decision mirror effect is broadly consistent with a resonance framework (Grossberg & Stone, 1986; Plaut et al., 1996; Van Orden & Goldinger, 1994). As noted earlier, the resonance framework proposes that word perception occurs in a series of cascaded stages (not by design but as an inevitable consequence of covariant learning), with each stage defined by the formation of stable feedback loops, or resonance. Because of tight statistical relations, O–P resonance occurs first, whereas O–S and P–S dynamics continue toward resonance (see Lukatela & Turvey, 1994a, 1994b; Perfetti & Bell, 1991; Perfetti & Zhang, 1995). When processing pseudohomophones, familiar patterns will emerge in O–P dynamics and P–S dynamics. Only a small misspelling exists to prevent O–S resonance, either by a verification process (Paap et al., 1982; Van Orden, 1987) or perhaps a *mismatch reset* function, as in adaptive resonance theory (Grossberg, 1980). Alternatively, people may be sensitive to disharmony in O–S dynamics, such that it provides a late "nonword" signal. By this view, words and pseudohomophones should produce exactly the profile we observed: The perceptual system cannot detect an O–S match or mismatch until all other processes have stabilized. Because this will always occur faster for more common patterns, performance to high-frequency words will exceed that to low-frequency words, and performance to high-frequency pseudohomophones will exceed that to low-frequency pseudohomophones.

REM-LD (Wagenmakers, Steyvers, et al., 2004). The missing element of the foregoing resonance account is a clearly defined decision stage. That is, the resonance framework can predict different time courses for emerging "word" and "nonword" signals, and is consistent with the lexical decision mirror effect. However, it cannot elegantly address global shifts in performance based on list composition, nonword difficulty, or the existence of associative priming relations among nonwords. Given the description from Van Orden and Goldinger (1994), the critical decision stage remains unspecified. The second theory that may explain our results is a perfect complement: *REM-LD* is a model of the decisional aspects of lexical decision, without any "front-end" perceptual mechanisms (Wagenmakers, Steyvers, et al., 2004). The REM-LD model is based on the memory model REM (retrieving effectively from memory; Shiffrin & Steyvers, 1997), which is easily extended to lexical decision. Various REM models have been described for different purposes, but all share basic elements: They describe how information is stored and retrieved from memory and how optimal (in this case, Bayesian) decisions can be made on the basis of noisy information.

As Wagenmakers, Steyvers, et al. (2004) explain, REM resembles many global memory models (e.g., MINERVA 2; Hintzman, 1986) by assuming that memory traces of high order units, such as words, consist of many lower level features. These features represent two broad classes of in-

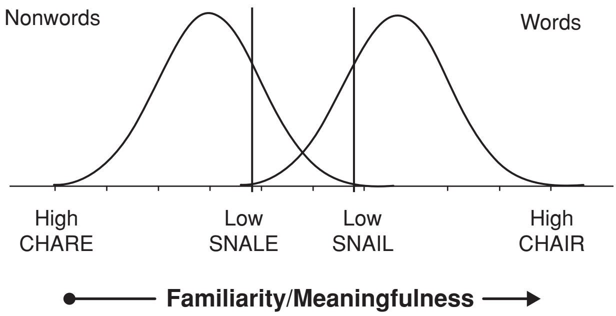

Figure 2. A lexical decision version of the mirror effect. Hypothetical nonword and word distributions are shown, with high-frequency items occupying the extremes.

formation: (1) properties of the words themselves (orthographic, phonologic, and semantic features) and (2) episodic, contextual information (aspects of the encoding event; Underwood, 1969). Although all traces are initially episodic, incorporation occurs over time. Thus, REM's architecture includes two broad classes of traces: *episodic* and *lexical-semantic* traces. Episodic traces contain noisy information about one specific encounter with a word. Lexical-semantic traces contain accumulated knowledge from many episodes, eventually producing a relatively complete and accurate trace (at least for the core features of each word). Therefore, the presentation of a known word has two effects. First, it creates a new episodic trace with noisy information about the item and about its context. Second, it updates the lexical-semantic trace, usually to a trivial degree. However, salient information, such as a surprising context or a unique font, can affect the lexical-semantic trace. In a memory task, such as recognition, people must use experimental context to discriminate "old" from "new" words. Thus, REM predicts that recognition or recall performance will reflect properties of the episodic (study) traces.

For more perceptual tasks, such as lexical decision, performance will mainly reflect properties of the lexical-semantic traces. In REM-LD, stimulus presentation creates a feature vector that is matched, in parallel fashion, to stored memory traces. Both probes and traces contain all features, but features become available gradually over time, partly on the basis of word frequency. As they become available, features in the probe and stored traces are evaluated, creating a set of matches and mismatches. Eventually, the system must generate a word/nonword decision; REM-LD assumes that maximum-likelihood decisions are generated, on the basis of prior and posterior odds. For simplicity, REM-LD generates a verification set for each probe item (see Paap et al., 1982) and then calculates matches, mismatches, and an odds ratio reflecting the likelihood that the stimulus is a real word. For the present purposes, an important aspect of REM-LD is that it can explicitly track the distributional properties that are implicit in SDT (see Gordon, 1983) and can thus predict effects of global context. Another important point is that "nonword" decisions are not reached by default but are computed in the exact manner specified for words. This creates a mechanism for the lexical decision mirror effect.

Taking these properties together, REM-LD portrays a decision stage that fits the results of Experiments 1A, 1B, 2A, and 2B. Indeed, the same mechanisms that explain the standard mirror effect in recognition (Shiffrin & Steyvers, 1997) are used to explain the reversed mirror effect we observed. The only underlying difference is that different subsets of memory traces naturally contribute to different experimental tasks. In fact, although they did not simulate the results, Wagenmakers, Steyvers, et al. (2004, p. 359) addressed the findings from Ziegler et al. (2001):

> It is worth mentioning one recent result with respect to the role of phonology in lexical decision: Ziegler, Jacobs, and Klüppel (2001) replicated in German results . . . showing that pseudohomophones derived from HF words are faster classified (i.e., correctly rejected) than pseudohomophones derived from LF words. . . . Such a result falls naturally out of REM models that incorporate differentiation. . . . The idea is that traces stored better are better differentiated from (i.e., less similar to) traces of other items. In the REM-LD model, we could assume that for both HF and LF pseudohomophones, their corresponding word is in the activated set. However, differentiation would mean that HF similarity would be lower than LF similarity. . . . To illustrate with an example, the well-stored information about BRAIN would produce relatively little confusion with BRANE, but the not-so-well stored information about FLOTSAM would produce relatively more confusion with FLOTSUM.

In short, REM-LD provides an explicit decision model, not unlike spelling verification in spirit, that could be naturally combined with a perceptual framework. We further consider the REM-LD model and its potential relation to the resonance framework in the General Discussion section. Next, we present Experiment 3, in which we focused on a possible connection between lexical decision and recognition memory.

EXPERIMENT 3

Presumably, given ample time, most pseudohomophones would be rejected in lexical decision, especially when people become vigilant to their presence (even in bottom-up theories, criteria are adjusted to reflect overall difficulty of discrimination). However, in typical lexical decision, people are encouraged to respond quickly. As illustrated by Van Orden and Goldinger (1994, Figure 6), lexical processing must proceed to full O–P–S (or global) resonance before discrimination of words and pseudohomophones can occur. The rate of achieving global resonance is jointly determined by ambiguity (cross talk) and frequency (see Plaut et al., 1996) and generally occurs sooner for higher frequency words. When performing lexical decision, people might wait for low-frequency pseudohomophones to reach their "breaking points," when O–S verification fails. Alternatively, they may false-alarm, which is potentially more interesting. In our method, when responses are entered, stimulus letter strings are immediately removed. According to the resonance view, in the absence of orthographic data, all ongoing processing (O–P resonance and burgeoning P–S and O–S dynamics) may conspire, creating the illusory perception of a correctly spelled word. We tested this prediction in Experiment 3, conducting a surprise recognition memory test after the participants completed the lexical decision task.

Method

Participants. Sixty-four students participated for course credit. All participants were native speakers of English with normal or corrected-to-normal vision.

Apparatus and Stimuli. All materials were those used in Experiments 1A, 1B, 2A, and 2B.

Procedure. The participants first completed two blocks of mixed-frequency lexical decision trials, identical to those in Experiment 2B. The only differences were that 128 trials were administered and feedback was withheld. On completing lexical decision, the participants received a surprise recognition test. All 128 base words were presented (without primes), all correctly spelled. The participants' task was to indicate whether each word was originally seen correctly or had been misspelled. Accuracy was encouraged over speed, although responses were required within 4 sec.

Results

All items were retained for analysis, but data from 3 participants were excluded for consistent failure to respond within the allotted time. The lexical decision results resembled those from Experiment 2B and are not considered further. The analyses focused on recognition: Hits and FAs were responses indicating that words and pseudohomophones, respectively, were originally experienced as proper words in lexical decision. The recognition data (raw probabilities) were analyzed in two sets, based on consistency with initial lexical decisions.

Table 4 shows recognition performance to words and pseudohomophones, including only items that generated consistent responses across tasks. For words, hit rates are shown, but only for words that were originally correctly classified (as words) in lexical decision. For pseudohomophones, FA rates are shown, including only items that were originally incorrectly classified (as words) in lexical decision. Thus, all results in Table 4 reflect the same underlying situation: The probability that letter strings were initially classified as "words" and were later remembered in kind (i.e., as having been correctly spelled during lexical decision).

Among the words, a frequency effect was observed [$F_1(1,60) = 17.23$, $p < .001$, $\eta_p^2 = .22$], with more hits for higher frequency words. Note that this result violates the standard pattern, wherein low-frequency words are advantaged in recognition memory. There was a reliable priming effect [$F_1(1,60) = 5.23$, $p < .05$, $\eta_p^2 = .08$], but this mainly reflected a frequency $\times$ priming interaction [$F_1(1,60) = 10.80$, $p < .01$, $\eta_p^2 = .15$]. The priming effect mainly occurred for low-frequency words [$t_1(60) = 3.33$, $p < .01$]. As shown in Table 4, hit rates to high-frequency words were unaffected by priming during initial exposure; priming improved hit rates to low-frequency words.

Among pseudohomophones, another frequency effect was observed [$F_1(1,60) = 27.67$, $p < .001$, $\eta_p^2 = .36$], with fewer FAs to high-frequency pseudohomophones. A priming effect was also observed [$F_1(1,60) = 6.57$, $p < .02$, $\eta_p^2 = .11$], with more FAs to pseudohomophones that were originally primed in lexical decision. The frequency $\times$ priming interaction was not reliable [$F_1(1,60) = 1.34$, $p > .2$]. As shown in Table 4, associative priming increased FAs to all pseudohomophones, regardless of base-word frequency.

Table 4
Mean Hits (to Original, Lexical Decision Words) and False Alarms (to Original, Lexical Decision Pseudohomophones) in Recognition Memory, Given Initial "Word" Responses in Lexical Decision

Prime Type	Low Frequency	High Frequency
Recognition Hits (Words)		
Unassociated	63.7 (80.4)	77.5 (81.2)
Associated	72.5 (78.2)	76.2 (80.6)
Recognition False Alarms (Pseudohomophones)		
Unassociated	5.5 (32.7)	2.4 (22.0)
Associated	8.3 (46.2)	3.5 (33.2)

Note—Values reflect raw percentages, followed by conditional percentages in parentheses (i.e., percentages of recognition hits and false alarms, from the set of items originally called "words" in lexical decision).

In terms of frequency effects, for both real words and those implied by pseudohomophones, the results in Table 4 present an intriguing pattern. As discussed earlier, the standard mirror effect in recognition memory occurs when low-frequency words yield more hits and fewer FAs, relative to high-frequency words (Glanzer et al., 1993). However, as shown in Table 4, a curious reversal of the effect was observed: Low-frequency words elicited fewer hits, and low-frequency pseudohomophones elicited more FAs. The result for real words is especially surprising: Even if the participants encoded their own lexical decision responses into memory for each word (see Logan, 1988), these words were correctly classified during encoding. We consider the pattern further after examining the complementary data.

Table 5 shows recognition performance to words and pseudohomophones, including only items that generated inconsistent responses across tasks. For words, these data are hit rates to words that were originally missed in lexical decision. For pseudohomophones, these data are FA rates for items that were originally correctly rejected in lexical decision. Thus, all results in Table 5 reflect the same underlying situation: letter strings that were initially classified as nonwords but were later remembered as correctly spelled words.

Considering words first, a large frequency effect was observed [$F_1(1,60) = 48.76$, $p < .001$, $\eta_p^2 = .45$], in the opposite direction to the results in Table 4: Now, low-frequency words elicited more hits than did high-frequency words. A reliable priming effect was observed [$F_1(1,60) = 27.29$, $p < .001$, $\eta_p^2 = .21$], also in the opposite direction to the results in Table 4: Hits were about half as likely for words that were primed during initial encoding. These factors interacted [$F_1(1,60) = 12.78$, $p < .001$, $\eta_p^2 = .18$], with stronger priming for low-frequency words.

Results for the pseudohomophones were also opposite to those observed in the consistent subset. As with words, there was a clear effect of original perception/response, but it reversed that seen with words: Having initially (correctly) rejected pseudohomophones, the participants were more likely to later false-alarm, remembering them as real words (note the inflated FA rates in Table 5, relative to those in Table 4). However, this comparison requires caution, as the base rates across tables are very uneven: Most pseudohomophones were correctly rejected in lexical decision, so more FAs were possible among the pseudohomophones listed in Table 5. This apparent effect was no longer evident when corrected, conditional probabilities were examined (shown in parentheses in Tables 4 and 5). The frequency effect was reliable [$F_1(1,60) = 98.10$, $p < .001$, $\eta_p^2 = .61$], again reversing the pattern in Table 4.

Table 5
Mean Hits (to Original, Lexical Decision Words) and False Alarms (to Original, Lexical Decision Pseudohomophones) in Recognition Memory, Given Initial "Nonword" Responses in Lexical Decision

Prime Type	Low Frequency	High Frequency
Recognition Hits (Words)		
Unassociated	12.5 (47.1)	2.9 (32.7)
Associated	5.2 (35.6)	1.5 (19.9)
Recognition False Alarms (Pseudohomophones)		
Unassociated	26.3 (31.1)	43.2 (45.9)
Associated	24.9 (30.8)	47.9 (33.2)

Note—Values reflect raw percentages, followed by conditional percentages in parentheses (i.e., percentages of recognition hits and false alarms, from the set of items originally called "nonwords" in lexical decision).

High-frequency pseudohomophones elicited nearly twice as many FAs as did low-frequency pseudohomophones. Priming created no reliable main effect or interaction.

Discussion

The results of Experiment 3 were partly anticipated and partly surprising. When low-frequency pseudohomophones, relative to high-frequency pseudohomophones, were incorrectly verified in lexical decision, they increased false recognition. This tendency was increased by associative priming during initial exposure. Given low-frequency pseudohomophones, we expected the participants to occasionally respond before achieving full global resonance, allowing perceptual dynamics to continue without the inhibiting influence of an incorrect letter string. We therefore predicted (and found) that, given FAs in lexical decision, the participants would be more likely to falsely recall seeing the actual low-frequency words.

A central aspect of the foregoing prediction was that FAs arise in lexical decision because people respond too early, before spelling verification failure (however conceived) can occur. Therefore, more surprising results came from pseudohomophones that were correctly rejected in lexical decision. By definition, correctly rejecting a pseudohomophone implies fairly complete processing, including the spelling verification. Also, one might expect participants to remember their own negative responses to those "words" (Logan, 1988; Whittlesea & Cantwell, 1987). Nevertheless, items that were correctly rejected in lexical decision still elicited high FA rates in recognition, whether assessed by raw or conditional probabilities. Moreover, the frequency effect was reversed, by a large degree, such that high-frequency pseudohomophones were more likely to elicit such FAs. Although we had not anticipated this pattern, it is broadly consistent with the phonologic coherence hypothesis. We consider this and other potential interpretations next.

GENERAL DISCUSSION

In the present experiments, we examined the perception and later memory of words and (more importantly) pseudohomophones, phonetically valid misspelled words, such as SNALE or NERSE. In lexical decision (Experiments 1A, 1B, 2A, and 2B), the observed responses to real words followed well-established patterns: Responses were faster to high-frequency words and were faster after associative priming. These factors interacted, with larger priming benefits to low-frequency words (Neely, 1991). The response patterns to pseudohomophones were more interesting, replicating and extending the results of Ziegler et al. (2001): In lexical decision, high-frequency pseudohomophones were privileged, leading to more efficient rejections than did low-frequency pseudohomophones. When considered together with the words, this suggested a lexical decision version of the mirror effect (Glanzer et al., 1993). As characterized in Figure 2, when participants must discriminate words from matched pseudohomophones, the stimuli are well characterized as reflected distributions, with the frequency-based extremes reversed, relative to the mirror effect in recognition memory.

As noted earlier, this outcome is inconsistent with the bottom-up similarity of nonwords to words and with predictions from bottom-up models of word perception (e.g., MROM and DRC). In most regards, high-frequency pseudohomophones should register high on the Balota and Chumbley (1984) F/M dimension. However, from the top down, people presumably have strong connections between high-frequency lexical representations and their spellings (Holmes & Ng, 1993; also Wagenmakers, Steyvers, et al., 2004). Thus, from the perspective of spelling verification, the expected organization of words and nonwords would follow our results: Higher frequency items should occupy the extremes, supporting fluent spelling verification or rejection. Stated differently, our results suggest that words and pseudohomophones self-organize into a configuration that optimizes top-down discrimination processes. Notably, the same result is predicted by REM-LD, a memory-based model of lexical decision, focused exclusively on the decision stage.

We previously suggested that these different explanations—the perception-based resonance account and the decision-based REM-LD account—should not be viewed as competing theories. Rather, they each characterize a stage of processing that is excluded by the other. The resonance framework is focused entirely on the dynamics that occur among knowledge sources during word perception, the journey from orthographic input to word perception. Word acceptance performance (especially latency) is well predicted by the resonance account, but accuracy and nonword rejection are mainly unspecified. Conversely, REM-LD simply presumes that all perceptual dynamics have occurred, such that an input probe is assembled in memory. For example, when discussing effects of pseudohomophones, Wagenmakers, Steyvers, et al. (2004, p. 357) wrote: "We assume that there are stages of processing that occur automatically en route to construction of the set of probe features, and that part of these stages involves production of phonological features." It seems likely that a resonance model could provide the missing perceptual stages, including the unfolding availability of features over time. As currently formulated, REM-LD does not predict RTs, although Wagenmakers, Steyvers, et al. described a diffusion process (as in Ratcliff, 1988) that would likely work. Because Experiments 1A, 1B, 2A, and 2B seem to have required a top-down process to predict RTs, we are interested to see this diffusion model implemented.

Two-Way Mirror Effects

Having replicated the lexical decision version of the mirror effect, we characterized top-down processing as *resonance building* (Grossberg & Stone, 1986; Plaut et al., 1996). By virtue of statistical properties, most resonance (typically connectionist) models naturally follow the phonologic coherence hypothesis (Van Orden & Goldinger, 1994). By this view, on presentation of a word (or pseudohomophone), O–P dynamics quickly achieve resonance, whereas O–S and P–S dynamics lag behind. With pseudohomophones, false-positive errors are prevented only by the failure of late-arriving O–S resonance. When people perform lexical decision under time pressure, FAs are inevitable. Low-frequency pseudohomophones, especially when presented in mixed-frequency blocks, seem most likely to elicit erroneous "word" responses prior to completed resonance.

From this perspective, we expected recognition errors in Experiment 3 to follow a specific pattern, with inflated FAs to low-frequency pseudohomophones. We reasoned that, given FAs in lexical decision, two forces would combine to encourage later false recognition: First, participants would encode their own (incorrect) responses (Logan, 1988; Whittlesea & Cantwell, 1987). Second, because pseudohomophones vanished on responding in lexical decision, lagging semantic-to-orthographic feedback may drive "hallucinations" of spellings that maximize self-consistency (see Grossberg, 1980). Indeed, half the results of Experiment 3 followed this pattern: Following FAs in lexical decision, low-frequency pseudohomophones elicited high-recognition FAs, relative to high-frequency pseudohomophones. However, another effect was also observed: Following correct rejections in lexical decision, high-frequency pseudohomophones elicited higher FA rates in recognition. Thus, the mirror effect had two faces, with diametrically opposite patterns based on initial lexical decision responses.

Although the latter effect was not anticipated, it allows (at least) two explanations, both relating to the relative coherence of perceptual processes for different words and nonwords. The potential accounts are similar to one another, but one emphasizes encoding processes, and the other emphasizes retrieval processes. On the encoding side, the result suggests that, despite being more efficiently rejected during lexical decision, high-frequency pseudohomophones implant stronger memories of their implied words. This is less contradictory than it may appear: By virtue of rapid O–P and P–S coherence, high-frequency pseudohomophones should strongly activate their implied words. Strong activation of familiar words allows fast, accurate detection of spelling errors (aberrant O–S correspondences; see earlier quote from Wagenmakers, Steyvers, et al., 2004). By way of analogy, an eyewitness with a clear image in mind will better reject foils in a lineup. However, this advantage in pseudohomophone rejection may create an ironic effect in later recognition: By strongly activating a word to reject its misspelling, a person may also create a strong memory that the actual word was encountered (see Jacoby, 1999, for a similar discussion).[1]

The retrieval-based account is somewhat similar, focusing on changes in perceptual fluency across the lexical decision and recognition procedures (Whittlesea & Leboe, 2000; Whittlesea & Williams, 2001). This is perhaps a more elegant account, based on two reasonable assumptions supported by prior data. First, assume that the perceptual processes (going from letter string to meaning) engaged by pseudohomophones are somewhat dysfluent, relative to those engaged by proper words. Second, assume that, among real words, perceptual processes are more fluent for more common words. Taking these together, the presentation of words during a recognition test will force an asymmetry: Whenever the original presentation is a pseudohomophone, test words (especially high-frequency test words) will elicit greater fluency, relative to the encoding event. As Whittlesea and colleagues (see also Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989) have shown, perceptual fluency often creates feelings of familiarity. Thus, another ironic effect will occur: By virtue of having first seen pseudohomophones, high-frequency words will elicit an unusually strong sense of perceptual fluency. Although this should alert people to a discrepancy between their past and present experiences, it has the opposite effect, triggering a sense of familiarity. In this regard, it is interesting to note that Wagenmakers, Zeelenberg, Steyvers, Shiffrin, and Raaijmakers (2004) accounted for (sometimes contradictory) nonword repetition effects by proposing similar dual processes: Repeated nonwords are harder to reject in lexical decision because their enhanced familiarity increases their "wordlikeness." But repeated nonwords are sometimes easy to reject because the participant remembers rejecting them previously. Using terms from Jacoby (1999), repetition creates both beneficial and ironic memory effects.

Note that, whichever of our accounts (if either) is correct, they share key elements: At heart, both presume that perceptual and memorial processes are tightly interconnected (Goldinger, Kleider, & Shelley, 1999; Roediger, 1996). Both also assume that degrees of self-consistency affect perceptual dynamics, leading to eventual differences in memory. Finally, in both lexical decision and recognition memory, there are shared assumptions that bottom-up dynamics are interpreted by top-down matching processes. The latter processes apparently define the topological relations among the experimental stimuli.

REFERENCES

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583-612). New York: Academic Press.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 340-357.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cogni-

tive effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 389-401.
Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.
Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, **121**, 275-277.
Forster, K. I., & Forster, J. C. (1996). DMASTR (version 2.16) [computer program]. Available at www.u.arizona.edu/~kforster/dmastr/dmastr.htm.
Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, **100**, 546-567.
Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, **18**, 381-398.
Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition*, **27**, 328-338.
Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning & Verbal Behavior*, **22**, 24-44.
Gottlob, L., Goldinger, S. D., Stone, G. O., & Van Orden, G. C. (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 561-574.
Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple readout model. *Psychological Review*, **103**, 518-565.
Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, **87**, 1-51.
Grossberg, S., & Stone, G. O. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, **93**, 46-74.
Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, **106**, 491-528.
Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.
Hintzman, D. L., & Curran, T. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General*, **126**, 228-247.
Holmes, V. M., & Ng, E. C. (1993). Word-specific knowledge, word-recognition strategies, and spelling ability. *Journal of Memory & Language*, **32**, 230-257.
Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 3-22.
Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.
Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, **118**, 126-135.
Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General*, **120**, 358-394.
Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, **122**, 316-330.
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.
Lukatela, G., Eaton, T., Lee, C. H., Carello, C., & Turvey, M. T. (2002). Equal homophonic priming with words and pseudohomophones. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 3-21.
Lukatela, G., Frost, S. J., & Turvey, M. T. (1999). Identity priming in English is compromised by phonological ambiguity. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 775-790.
Lukatela, G., & Turvey, M. T. (1994a). Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, **123**, 107-128.
Lukatela, G., & Turvey, M. T. (1994b). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General*, **123**, 331-353.
Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86**, 287-330.
Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.
Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review*, **89**, 573-594.
Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 msec of word identification: Evidence from backward masking and masked priming. *Journal of Memory & Language*, **30**, 473-485.
Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 24-33.
Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.
Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review*, **95**, 238-255.
Rhodes, G., Parkin, A. J., & Tremewan, T. (1993). Semantic priming and sensitivity in lexical decision. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 154-165.
Roediger, H. L., III (1996). Memory illusions. *Journal of Memory & Language*, **35**, 76-100.
Rueckl, J. G. (2002). The dynamics of visual word recognition. *Ecological Psychology*, **14**, 5-19.
Schvaneveldt, R. W., & McDonald, J. E. (1981). Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 673-687.
Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, **96**, 523-568.
Seidenberg, M. S., & Zevin, J. D. (2003, November). *Sensitivity to sublexical statistics and subject variability in the triangle model of reading*. Poster presented at the annual meeting of the Psychonomic Society, Vancouver, BC, Canada.
Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.
Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory & Language*, **36**, 337-359.
Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 744-774.
Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, **76**, 559-573.
Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, **15**, 181-198.
Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains the perception of printed words. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1269-1291.
Van Orden, G. C., Stone, G. O., Garlington, K. L., Markson, L. R.,

Pinnt, G. S., Simonfy, C. M., & Brichetto, T. (1992). "Assembled" phonology and reading: A case study in how theoretical perspective shapes empirical investigation. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 249-292). Amsterdam: North-Holland.

Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology*, **48**, 332-367.

Wagenmakers, E.-J., Zeelenberg, R., Steyvers, M., Shiffrin, R. M., & Raaijmakers, J. G. W. (2004). Nonword repetition in lexical decision: Support for two opposing processes. *Quarterly Journal of Experiment Psychology*, **57A**, 1191-1210.

Whittlesea, B. W. A., & Cantwell, A. L. (1987). Enduring influence of the purpose of experiences: Encoding–retrieval interactions in word and pseudoword perception. *Memory & Cognition*, **15**, 465-472.

Whittlesea, B. W. A., & Leboe, J. (2000). The heuristic basis of remembering and classification: Fluency, generation and resemblance. *Journal of Experimental Psychology: General*, **129**, 84-106.

Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 3-13.

Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 547-559.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers*, **29**, 600-618.

NOTE

1. One limitation of Experiment 3 is that all encoding was performed in a speeded, incidental learning task. It will be important to replicate these patterns of true and false recognition in conditions without speeded encoding and without the surprise recognition test.

APPENDIX A
Stimulus Materials

Word	Nonword	Prime
	Low-Frequency Items	
peach	peech	apricot
groom	grume	bride
cheer	chier	hurrah
thorn	thourn	rosebush
broom	brume	dust-pan
nail	nale	hammer
fright	frite	terror
soak	soke	bathe
grease	greese	lubricate
comb	coam	brush
bleed	blead	wound
swarm	sworm	bees
haste	haist	hurry
grief	greaf	despair
cloak	cloke	cape
jerk	jirk	yank
wheat	wheet	grain
fern	fearn	moss
stance	stanse	position
chord	coard	rope
trance	transe	hypnosis
grope	groap	fondle
stroll	stroal	walk
bleak	bleek	dismal
hurl	hirl	throw
bruise	brooze	welt
meek	miek	mild
smirk	smurk	grimace
sparse	sparce	scattered
frail	frale	weak
bloat	blote	swell
dread	dred	worry
thumb	thum	finger
pants	pance	shirt
freak	freek	oddball
sponge	spunge	mop
spoon	spune	fork
tease	teeze	taunt
blaze	blaise	flames

(*Continued on next page*)

APPENDIX A (Continued)

Word	Nonword	Prime
freeze	freize	chill
jade	jaid	emerald
cheat	cheet	lie
cruise	crooze	vacation
bead	beed	necklace
soar	soor	fly
weird	weard	strange
stool	stule	chair
seize	seeze	confiscate
creep	creap	crawl
gleam	gleem	glisten
chess	ches	checkers
weave	weeve	knit
niece	neace	nephew
brute	brewt	bully
wreath	wreeth	garland
kneel	kneal	squat
rinse	rince	wash
peep	peap	chirp
hose	hoze	fire-engine
crease	criece	wrinkle
whirl	whurl	spin
toad	tode	frog
poke	poak	jab
yearn	yurn	longing
	High-Frequency Items	
date	dait	time
heat	heet	cold
front	frunt	back
same	saim	different
need	nead	want
class	clas	teacher
year	yeer	annual
done	dun	finished
earth	erth	planet
game	gaim	play
home	hoam	house
plane	pleign	jet
wife	whife	husband
force	forse	brute
note	noat	letter
small	smaul	big
death	deth	life
speak	speek	talk
east	eest	west
green	grean	grass
growth	groath	development
read	rede	book
trade	traid	exchange
close	cloze	shut
hear	heer	listen
near	neer	close
group	groop	gang
serve	surve	tennis
fight	fite	brawl
rate	rait	heart
first	furst	last
weeks	weaks	days
deep	deap	shallow
dark	darc	shadow

APPENDIX A (Continued)

Word	Nonword	Prime
gas	gass	oil
words	werds	sentence
late	lait	early
church	cherch	religion
blue	bloo	sky
peace	peece	war
paid	pade	bought
work	werk	job
room	rume	dorm
hope	hoap	wish
white	wite	snow
court	cort	trial
fear	feer	scary
firm	ferm	hard
reach	reech	grasp
lead	leed	follow
leave	lieve	depart
feel	feal	touch
from	frum	to
girl	gurl	boy
street	strete	road
case	cace	detective
board	bord	plank
wait	wate	pause
four	foar	three
mean	mene	cruel
young	yung	old
great	grait	super
least	liest	most
corps	coar	marine

(*Continued on next page*)

APPENDIX B
Statistical Results for Experiments 1A and 1B

Experiment 1A			Experiment 1B		
F	p	η_p^2	F	p	η_p^2
Hit Response Times, Words: 2 × 2 ANOVAs Across Priming (Unrelated, Related) and Frequency (High, Low)					
Priming					
$F_1(1,59) = 8.7$	$<.01$	.12	$F_1(1,27) = 13.9$	$<.01$	.34
$F_2(1,113) = 24.8$	$<.001$	.18	$F_2(1,113) = 10.7$	$<.01$	.09
Frequency					
$F_1(1,59) = 30.4$	$<.001$	.34	$F_1(1,27) = 58.9$	$<.001$	.67
$F_2(1,113) = 54.5$	$<.001$	.33	$F_2(1,113) = 44.2$	$<.001$	.28
Priming × Frequency					
$F_1(1,59) = 1.8$	n.s.	n.a.	$F_1(1,27) = 7.1$	$<.05$	.05
$F_2(1,113) = 1.4$	n.s.	n.a.	$F_2(1,113) = 6.5$	$<.05$	.06
One-Way ANOVA: Frequency Effects in Pseudohomophones					
Frequency					
$F_1(1,59) = 2.9$	n.s.	n.a.	$F_1(1,27) = 8.7$	$<.01$	.24
$F_2(1,113) = 11.9$	$<.01$	.10	$F_2(1,113) = 3.8$	.053	.03
Miss Rates, Words: 2 × 2 ANOVAs Across Priming (Unrelated, Related) and Frequency (High, Low)					
Priming					
$F_1(1,59) = 5.4$	$<.05$	.08	$F_1(1,27) = 18.1$	$<.001$	.40
$F_2(1,113) = 10.7$	$<.001$	.09	$F_2(1,113) = 22.1$	$<.001$	.16
Frequency					
$F_1(1,59) = 24.5$	$<.001$	.29	$F_1(1,27) = 8.5$	$<.01$	.23
$F_2(1,113) = 28.2$	$<.001$	.20	$F_2(1,113) = 12.2$	$<.01$	.10
Priming × Frequency					
$F_1(1,59) < 1.0$	n.s.	n.a.	$F_1(1,27) = 3.9$	.06	.13
$F_2(1,113) = 1.4$	n.s.	n.a.	$F_2(1,113) = 9.0$	$<.01$	.07
One-Way ANOVA: Frequency Effects in Pseudohomophones					
Frequency					
$F_1(1,59) = 15.1$	$<.001$	.20	$F_1(1,27) = 20.0$	$<.001$	.43
$F_2(1,113) = 17.4$	$<.001$	.14	$F_2(1,113) = 10.9$	$<.01$	.09

Note—F_1 denotes participants analyses; F_2 denotes items analyses.

APPENDIX C
Statistical Results for Experiments 2A and 2B

Experiment 2A			Experiment 2B		
F	p	η_p^2	F	p	η_p^2
Hit Response Times, Words: 2 × 2 ANOVAs Across Priming (Unrelated, Related) and Frequency (High, Low)					
Priming					
$F_1(1,56) = 28.4$	<.001	.34	$F_1(1,30) = 8.9$	<.001	.39
$F_2(1,113) = 27.6$	<.001	.20	$F_2(1,113) = 22.2$	<.001	.33
Frequency					
$F_1(1,56) = 11.8$	<.01	.17	$F_1(1,30) = 100.3$	<.001	.77
$F_2(1,113) = 68.8$	<.001	.33	$F_2(1,113) = 64.5$	<.001	.36
Priming × Frequency					
$F_1(1,56) < 1.0$	n.s.	n.a.	$F_1(1,30) = 4.6$	<.05	.04
$F_2(1,113) < 1.0$	n.s.	n.a.	$F_2(1,113) = 3.7$	.057	.03
Miss Rates, Words: 2 × 2 ANOVAs Across Priming (Unrelated, Related) and Frequency (High, Low)					
Priming					
$F_1(1,56) = 16.3$	<.001	.23	$F_1(1,30) = 18.0$	<.001	.49
$F_2(1,113) = 16.2$	<.001	.13	$F_2(1,113) = 15.0$	<.001	.16
Frequency					
$F_1(1,56) = 18.7$	<.001	.25	$F_1(1,30) = 28.7$	<.001	.47
$F_2(1,113) = 21.4$	<.001	.16	$F_2(1,113) = 36.2$	<.001	.24
Priming × Frequency					
$F_1(1,56) = 5.3$	<.05	.09	$F_1(1,30) = 4.5$	<.05	.13
$F_2(1,113) = 4.7$	<.05	.04	$F_2(1,113) = 5.4$	<.05	.05
Correct Rejection Response Times, Pseudohomophones: 2 × 2 ANOVAs Across Priming (Unrelated, Related) and Frequency (High, Low)					
Priming					
$F_1(1,56) < 1.0$	n.s.	n.a.	$F_1(1,30) < 1.0$	n.s.	n.a.
$F_2(1,113) < 1.0$	n.s.	n.a.	$F_2(1,113) < 1.0$	n.s.	n.a.
Frequency					
$F_1(1,56) = 12.4$	<.01	.18	$F_1(1,30) = 7.2$	<.05	.19
$F_2(1,113) = 67.8$	<.001	.38	$F_2(1,113) = 7.8$	<.01	.07
Priming × Frequency					
$F_1(1,56) = 5.7$	<.05	.09	$F_1(1,30) = 2.3$	n.s.	n.a.
$F_2(1,113) = 67.8$	<.001	.38	$F_2(1,113) = 2.6$	n.s.	n.a.
False-Alarm Rates, Pseudohomophones: 2 × 2 ANOVAs Across Priming (Unrelated, Related) and Frequency (High, Low)					
Priming					
$F_1(1,56) = 2.8$	n.s.	n.a.	$F_1(1,30) = 4.5$	<.05	.13
$F_2(1,113) = 1.7$	n.s.	n.a.	$F_2(1,113) = 5.9$	<.05	.05
Frequency					
$F_1(1,56) = 21.3$	<.001	.28	$F_1(1,30) = 42.5$	<.001	.59
$F_2(1,113) = 17.3$	<.001	.13	$F_2(1,113) = 13.3$	<.001	.11
Priming × Frequency					
$F_1(1,56) = 7.4$	<.01	.12	$F_1(1,30) = 2.9$	n.s.	n.a.
$F_2(1,113) = 5.1$	<.05	.04	$F_2(1,113) = 5.0$	<.05	.04

Note— F_1 denotes participants analyses; F_2 denotes items analyses.

(Manuscript received July 19, 2004;
revision accepted for publication June 28, 2005.)