Paws + cause = pause? Memory load and memory blends in homophone recognition

TAMIKO AZUMA, ERICA J. WILLIAMS, and JULIET E. DAVIE
Arizona State University, Tempe, Arizona

Previous research has shown that separate information sources, activated closely together in time, can induce errors suggestive of memory blends. In Experiment 1, homophones were used to induce such memory errors. In a study task, participants made relatedness judgments to word pairs that included homophones (e.g., PAWS–BEAR). During this task, one group (study-similar) maintained memory loads with words that were orthographically similar to the presented homophones (e.g., JAWS). Another group (test-similar) maintained memory loads similar to those obtained for the study homophone's alternate spelling (e.g., CAUSE). A third group maintained no memory load during the task. In a surprise recognition test, participants were presented both previously viewed homophones (PAWS) and non-presented alternate spellings (PAUSE). We hypothesized that partially activated alternate spellings, in conjunction with the orthographically similar memory words, would result in the creation of blended memories. The results followed suit: The test-similar condition produced significantly elevated false alarms, relative to both the study-similar and the no-load conditions. Experiment 2 replicated these results while including memory loads to control for potential orthographic confusions. The findings are discussed in terms of multiple, episodic memories later activated as single, blended memories.

A basic assumption of almost all memory models, including multiple-trace (Hintzman, 1986; Shiffrin & Steyvers, 1997) and spreading-activation models (Collins & Loftus, 1975; McClelland & Rumelhart, 1985), is that related items activate each another. For example, in spreading-activation models, perception of a target word results in concurrent activation of surrounding related words. More recent models propose that, with sufficient similarity, information from separate events may be retrieved together, causing *memory blends*, “memory” for nonexperienced stimuli or events (e.g., Arndt & Hirshman, 1998; Stark & McClelland, 2000).

False memory errors have been frequently observed in the context of single words. Using the now popular Deese (1959), or DRM, paradigm, Roediger and McDermott (1995, 2000) had people study semantically related words (e.g., *bed, rest, awake*, etc.), all related to a critical non-presented word (e.g., *sleep*). Afterward, participants falsely recalled such critical words with the same frequency as they did legitimate targets. A natural account assumes that study words repeatedly activate such critical lures. During recall, people erroneously attribute the previous activation of critical words to their inclusion in study

lists. Such findings exist for words that are similar along other, nonsemantic dimensions, such as orthography (Pesta, Murphy, & Sanders, 2001) or phonology (Sommers & Lewis, 1999). Recently, Watson, Balota, and Roediger (2003) presented lists containing both semantic and phonological associates. For those hybrid lists, false recall of critical lures exhibited over-additivity, exceeding the level expected from the simple sum of errors produced by separate semantic and phonological lists. Watson et al. interpreted these results in terms of separate lexical dimensions combining into potent false memories.

Depending on the original encoding task, words may be represented in memory with emphasis on their semantic, orthographic (visual), or phonological dimensions. In a recognition test, all three dimensions may be evaluated. In distributed models, words may be represented as convolved patterns of activation across these dimensions. If words are similar or are related along one dimension, it is assumed that their representations overlap. Thus, multiple words may be activated via shared units. For example, research on homophones (words with different spellings but shared pronunciation, such as PAIN and PANE) has produced findings indicating the activation of multiple orthographic representations through shared phonological units. Homophones cause many errors (and/or slow response times) in simple language tasks, such as judging relatedness (Jared & Seidenberg, 1991; Luo, Johnson, & Gallo, 1998). For example, Van Orden (1987) observed high false-positive rates to homophones in semantic categorization (e.g., verifying “hare” as “part of the body”). In addition, lexical decisions are slow and inaccurate to homophones, relative to non-

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R03-DC04231-03 to T.A. Special thanks to Steve Goldinger for valuable suggestions and to Krista Matsueda and Trisha Palma for making this research possible. We are also grateful to Michael Humphreys, Jason Arndt, and Henry Roediger III for clear and helpful comments. Correspondence should be addressed to T. Azuma, Department of Speech and Hearing Science, P. O. Box 870102, Arizona State University, Tempe, AZ 85287-0102 (e-mail: azuma@asu.edu).

homophones (Pexman, Lupker, & Jared, 2001). Some researchers theorize that homophone effects reflect the activation of multiple orthographic representations through shared phonological units; competition between these orthographic patterns slows responses (Pexman et al., 2001; Stone, Vanhoy, & Van Orden, 1997; Ziegler, Montant, & Jacobs, 1997). Lukatela, Lukatela, and Turvey (1993; also Lukatela & Turvey, 1994) found that homophones (e.g., TOWED) primed associates of their alternative spellings (e.g., FROG). This "false priming" effect arises very early in word perception; it is observed with brief stimulus onset asynchronies (SOAs), vanishing with longer SOAs. Similarly, Rueckl and Mathew (1999) found that visual presentation of homophones results in priming for both spellings in a subsequent stem completion task. They proposed that "the outcome of a processing event reflects the contribution of many operations, each of which leaves a memory trace that will, under the appropriate conditions, influence subsequent behavior. In the case of visual word identification, these processes include the construction of orthographic, phonological, and semantic codes" (p. 9).

Distributed models of lexical access (e.g., Van Orden, 1987; Van Orden & Goldinger, 1994) assume that presentation of a homophone activates both spellings through shared phonology (although nonpresented homophones will be activated to lesser degrees). Given this assumption, it should be possible to induce memory errors by using words with shared phonology similar to those observed for words with overlapping semantic representations. Humphreys, Burt, and Lawrence (2001) examined blended memory errors for *orthographically* similar words. In their study, participants viewed primes and named targets. In the critical condition, semantic associations were highly expected, due to relatedness proportion. Also, people were placed under time pressure in order to implicitly encourage strategic anticipation of targets. Thus, given the prime *soil*, participants should strongly activate the semantic associate *dirt*. However, some related targets were replaced with orthographically similar words (e.g., *dart*). As a result, the expected word and the orthographically similar target should be activated almost simultaneously, one from the top down, the other from the bottom up. Later, participants received a recognition test that included the actual targets, unprimed foils, and primed (expected) foils. Humphreys et al. found that presentation of the orthographically similar targets (*dart*) increased false alarms to the expected words (*dirt*). They proposed that such memory errors were due to the creation of blended episodic memories, created from "two different semantic entities."

In the present research, people saw homophones in a study task in which they judged word pairs as related or unrelated. They later received a surprise recognition test, including both the study homophones and their alternate spellings. During the study task, some participants maintained memory loads. In the *study-similar* condition, people remembered words that were orthographically sim-

ilar to the study homophones. For example, if *HERE* and *PAWS* were presented during study, the memory load included the words *MERE* and *JAWS*. In the *test-similar* condition, people remembered words that were orthographically similar to the alternate (nonpresented) spellings of the study homophones. For example, if *HERE* and *PAWS* were presented at study, the memory load included *NEAR* (similar to *HEAR*) and *CAUSE* (similar to *PAUSE*). In the surprise recognition test, participants made judgments to both spellings of the homophones (e.g., *PAWS* and *PAUSE*).

The following predictions were based on the assumptions of distributed models of memory and lexical access, and previous findings of memory errors with semantically or orthographically similar words. The crucial prediction concerned the test-similar condition: Given the content of the memory load, we predicted that this condition would increase false alarms, relative to other conditions. For example, presentation of the study homophone *PAWS* should partially activate *PAUSE*. In the test-similar condition, participants must actively maintain an orthographically similar word (*CAUSE*). Like Humphreys et al. (2001), we expected that simultaneous activation of the nonpresented homophone (*PAUSE*) and the orthographically similar word (*CAUSE*) would lead to blended memories. If so, people would be more likely to false alarm to the never presented alternate spellings of the homophones.

The study-similar and no-load conditions were used to determine whether maintaining *any* memory load during study would increase false alarms to the homophones' alternative spellings. If so, the test-similar and study-similar conditions would result in equivalent error rates, both exceeding that for the no-load condition. In the study-similar condition, it was also possible that the orthographic similarity of the memory load words to the study homophones would actually improve hit rates. That is, the memory words could provide additional activation for study homophones via overlapping orthographic representations.

EXPERIMENT 1

Method

Participants. One hundred twenty-four introductory psychology students at Arizona State University received partial course credit for their participation. All participants spoke English as their primary language.

Stimuli. Two word lists were created for the study task (relatedness judgment). Both lists contained 18 homophone pairs and 18 unrelated word pairs. Each homophone pair included a homophone and a semantic associate (e.g., *PAWS*–*BEAR*). Spelling dominance of the homophones was counterbalanced across lists: In each list, nine homophones were tested in their dominant spellings and nine in their subordinate spellings, as determined with published familiarity ratings (Kreuz, 1987). Each list contained only one spelling of a given homophone.

Two different sets of memory loads were created for each list. The study-similar load contained words that were orthographically similar to the study homophones (e.g., study homophone–*PAWS*; memory load word–*jaws*). The test-similar load contained words that were orthographically similar to the alternate spellings of the

study homophones (e.g., study homophone–PAWS; memory load word–*cause*). With one exception (*peace–cease*), each memory load word was only one letter (or consonant cluster) different from its matched homophone. In summary, there were three different memory load conditions (no-load, study-similar, and test-similar), each with two different lists of study homophone pairs. All conditions were followed by the same recognition test which included the 18 study homophones and their alternative spellings (for a total of 36 trials).

Procedure. Participants were run in groups of 1 to 8 on PC-compatible computers with standard monitors. Forty individuals were randomly assigned to each of three memory load conditions. In the study task, participants made relatedness judgments to word pairs (e.g., PAWS–BEAR) displayed on the computer screen, quickly pressing “yes” or “no” keys on the keyboard. In all conditions, participants performed three blocks of 12 relatedness trials (6 related, 6 unrelated), for a total of 36 pairs. In the memory load conditions, participants tried to remember word lists during the study task. Before each block, they were given 20 sec to learn six words. Subsequently, the words were removed, and the relatedness task began. After the relatedness trials, participants were prompted to write down all possible memory load words. In the no-load condition, people were shown a blank screen for 20 sec in place of the memory load words. After all three study blocks, all participants completed the same distractor task, consisting of 15 arithmetic problems shown on the computer.

Following the distractor task, participants were given a surprise recognition test. The same recognition test was used for everyone, containing both spellings of all 18 previous homophones. Participants were told to decide whether each word had been shown in the previous relatedness task. They were explicitly warned that the words had multiple spellings, and they were instructed to respond “old” only to correct spellings. The homophones were displayed one at a time, in random order, and they remained on the screen until the participant responded “yes” or “no” by pressing designated keys.

Results

Four participants either failed to achieve 80% accuracy in the relatedness task or consistently recalled fewer than two memory load words. These individuals were dropped and replaced with other participants. One-way analyses of variance (ANOVAs) were used to evaluate the effect of memory load condition (no-load, study-similar, test-similar) on performance. Planned comparisons among the memory load conditions were assessed using Fisher’s LSD tests (with $\alpha = .05$).

Mean response times (RTs) and accuracy rates for the relatedness judgments (study task) are presented in Table 1. Analyses revealed no effect of memory load on relatedness judgment times [$F(2,117) = 1.46, p = .24$] or ac-

Table 1
Mean Correct Response Times (RTs, in Milliseconds) and Percent Accuracy, With Standard Deviations, on the Study (Relatedness Judgment) Task in Experiment 1

Memory Load Condition	Correct RTs		% Accuracy	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No load	1,836	697	93.7	5.2
Study similar	2,096	782	96.3	6.5
Test similar	1,949	544	95.2	6.5

curacy [$F(2,117) = 1.82, p = .17$]. Thus, maintenance of memory loads did not strongly affect performance in the study task. Results for the recognition test, shown in Table 2, include mean adjusted recognition rates, correct RTs, hit rates, false alarm rates, and nonparametric signal detection indexes. (Adjusted recognition rates were calculated by subtracting each participant’s false alarm rate from his or her hit rate.) There was a significant main effect of memory load [$F(2,117) = 12.02, p < .001$]. Planned comparisons revealed significantly higher recognition accuracy in the no-load condition, relative to both memory load conditions. This pattern was replicated in correct RTs, with another significant main effect of memory load [$F(2,117) = 5.90, p < .01$]. Post hoc comparisons verified that responses were faster in the no-load condition than in the study-similar condition and were marginally faster ($p = .05$) than in the test-similar condition. The effect of memory load on hit rate was similar [$F(2,117) = 13.46, p < .001$], with the no-load condition having a significantly higher hit rate than those for both memory load conditions. The study-similar and test-similar conditions did not differ from each other. These findings all illustrate the detrimental impact of maintaining a memory load during study.

The most interesting result involved the significant effect of memory load on false alarms [$F(2,117) = 5.48, p < .01$]. Planned comparisons revealed that the test-similar condition produced more false alarms than did the no-load and study-similar conditions (which did not differ from each other). Nonparametric signal detection indexes of sensitivity (A') and bias (B''_D), were also calculated for each participant (Donaldson, 1992). A main effect of memory load on A' was observed [$F(2,117) = 11.66, p < .01$], with higher A' in the no-load condition than in either memory load condition. For B''_D , one-sample

Table 2
Mean Recognition Test Performance and Signal Detection Variables for the Three Memory Load Conditions in Experiment 1

Memory Load	Adjusted Recognition Rate		Correct RTs (in msec)		Hit Rate		False Alarm Rate		A'		Criterion (B''_D)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No load	82.4	13	1,593	452	93.3	7	10.9	10	0.95	.04	−0.18 [‡]	0.67
Study similar	65.8*	20	1,945*	496	78.0*	17	12.2	12	0.89*	.07	0.18 [‡]	0.58
Test similar	64.3*	21	1,794*	478	82.4*	14	18.1 [†]	11	0.88*	.08	−0.01 [‡]	0.53

*Significantly different from no-load condition at the .05 level (Fisher’s LSD test). [†]Significantly different from no-load and study-similar conditions at the .05 level (Fisher’s LSD test). [‡]Not significantly different from zero (i.e., no evidence of bias) with one-sample *t* test.

t tests revealed that no condition differed significantly from zero (i.e., no evidence of significant bias in either direction). In a post hoc analysis, we also assessed whether false alarm rates in the test-similar condition were dependent on the presentation order of the homophones (presented vs. new) during the recognition test. Although false alarm rates were slightly lower for first encounters (.16) relative to second encounters (.20), the difference was not significant [$F(1,38) = 1.45, p = .24$].

To examine the possible relationship between individual working memory (WM) capacity and recognition performance, the results of the memory load task were analyzed. It was assumed that superior memory load performance reflected higher WM capacity. The total number of correctly recalled memory load words was calculated for each individual in both load conditions. The test-similar and study-similar conditions did not significantly differ ($M = 13.28$ and 13.03 , respectively). Memory load performance was then correlated with the measures of recognition test performance (see Table 3). Both memory load conditions showed significant positive correlations between memory load and recognition performance (including A'). This result seems to verify that memory load performance reflects some basic aspect of processing, such as WM capacity. For the test-similar condition, the most striking result was a strong relationship between memory load performance and false alarm rate ($r = -.59$). As WM capacity decreased, false alarms to alternate spellings in the recognition test increased. In the study-similar condition, there was also a significant correlation between memory load performance and false alarm rate, but the relationship was weaker ($r = -.37$).

EXPERIMENT 2

The results of Experiment 1 suggest that people make blend memory errors when simultaneously presented with homophones (e.g., PAWS) and memory load words that resemble their alternate spellings (e.g., CAUSE). However, two aspects of the design limited its interpretation. First, the three load conditions were presented to different groups of participants. Thus, the results may have reflected differences across groups (e.g., different response criteria). In Experiment 2, the memory load conditions were changed to a within-subjects variable.

Second, rather than demonstrating blend memories, the high false alarm rate in the test-similar condition may have simply reflected the requirement to memorize load words that closely resembled later test words. That is, memorizing CAUSE could make people falsely recognize PAUSE at test, even if PAWS was never encountered in the study list. To test this alternative account, a control condition was needed wherein participants memorized CAUSE and were later tested on PAUSE, but the study word PAWS was absent. If this condition still induced high false alarm rates, it would suggest that simple resemblance, rather than memory blending, created the pattern in Experiment 1. Thus, in Experiment 2, the no-load condition was replaced with an orthographic control condition in which the memory load words were orthographically similar to test words, but not to any study words.

Method

Participants. Fifty-one introductory psychology students at Arizona State University received partial course credit for participation. All participants spoke English as their primary language.

Stimuli. Three lists of word pairs and memory loads were created for the study task. Each list had corresponding study-similar, test-similar, and orthographic control loads. The study-similar and test-similar loads from Experiment 1 were used again. The orthographic control load contained words that were similar to homophones presented in the recognition test, but not to any study words. For example, given the test word PAUSE, the orthographic control memory load word was CAUSE, but the study homophone PAWS was replaced with SENT. Assignment of words to conditions was counterbalanced; conditions were completed in random order.

The recognition test was composed of 48 words (24 old, 24 new). These included 12 old homophones (from the study-similar and test-similar conditions), 6 old homophones that had no counterparts from the memory loads, and 6 old nonhomophonic words. These were combined with 24 new homophones: Twelve were alternate spellings of the study-similar and test-similar homophones (6 each), 6 were new homophones resembling the orthographic control words, and 6 were their alternate spellings.

Procedure. The procedure was the same as used in Experiment 1.

Results

Four participants either failed to achieve 80% accuracy in the relatedness task or consistently recalled fewer than two memory load words. These individuals were dropped and replaced with other participants. The recognition data for each condition are shown in Table 4, including mean adjusted recognition rates, correct RTs, hit

Table 3
Performance on Memory Load (ML) Recall and Correlations With
Recognition Test Variables in Experiment 1

	Study Similar Condition	Test Similar Condition
Mean ML performance	13.03 (2.41)	13.28 (2.42)
Correlation of ML $\times$ false alarm rate	-.372*	-.591‡
Correlation of ML $\times$ hit rate	-.223	.192
Correlation of ML $\times$ recognition rate	.358*	.454†
Correlation of ML $\times$ A'	.372*	.430†
Correlation of ML $\times$ criterion (B''_d)	.020	.308*

*Correlation significant at the .05 level (two-tailed). †Correlation significant at the .01 level (two-tailed). ‡Correlation significant at the .001 level (two-tailed).

Table 4
Mean Recognition Test Performance and Signal Detection Variables for Experiment 2

Memory Load	Adjusted Recognition Rate		Correct RTs (in msec)		Hit Rate		False Alarm Rate		A'		Criterion (B''_D)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Orthographic control	N/A		1,644	654	N/A		18.4*	17	N/A		N/A	
Study similar	58.3	33	1,567	654	78.1	16	19.7*	20	0.85	.14	-0.24†	.65
Test similar	57.4	28	1,616	533	82.6	18	25.2	19	0.83	.18	0.25†	.19

*Significantly different from test-similar condition ($p < .05$). †Not significantly different from zero with one-sample t test.

rates, false alarm rates, and nonparametric signal detection indexes. As in Experiment 1, the study-similar and test-similar conditions did not significantly differ in adjusted recognition rates ($F < 1$), correct RTs ($F < 1$), hit rates [$F(2,100) = 2.34$], or A' [$F(2,100) = 1.27$]. None of the criteria significantly differed from zero (all one-sample t tests). Most important, the test-similar condition again produced significantly more false alarms than did the study-similar and the orthographic control conditions [$F(2,100) = 3.34$, $p < .05$]. In fact, the false alarm rate was numerically smallest for the critical orthographic control condition.

As in Experiment 1, correlational analyses were conducted to examine the potential relationship between individual WM capacity, estimated by memory load performance, and measures of recognition performance. Surprisingly, WM was not significantly correlated with any recognition measures, although the numeric trends were similar to those in Experiment 1. Examination of the raw data revealed restricted ranges for several recognition variables, particularly among participants with higher WM. Using a median split, we classified participants into *high-* or *low-WM* groups and conducted separate correlations. For the high-WM group, memory load was not significantly correlated with any recognition measures, including false alarm rate [$r(27) = -.257$, $p = .20$]. However, for the low-WM group, memory load was significantly correlated with false alarm rate [$r(24) = -.420$, $p < .05$]. Although the directions of association replicated those of Experiment 1, the restricted ranges preclude clear interpretation.

DISCUSSION

In Experiment 1, overall analyses showed that maintaining any WM load during study impaired subsequent recognition performance, relative to a no-load condition. This finding is not surprising; participants who devote more attentional resources to studying the homophones should have stronger memory for those words. More important, the test-similar condition increased false alarms, relative to the other conditions. People were more likely to erroneously “recognize” nonpresented homophones when orthographically similar words were included in the memory load. The no-load and study-similar conditions did not differ in false alarms, suggesting that the *nature* of the memory load affects false recognition.

There was also a strong negative relationship between individual WM capacity, estimated by memory load performance, and test-similar false alarm rates. Experiment 2 replicated the results in a within-subjects design. In addition, the orthographic control data showed that memory load words do not induce high false alarm rates when the potential for blending has been removed.

The present findings complement those of Humphreys et al. (2001), who reported evidence of memory blends when self-generated words and orthographically similar targets were simultaneously experienced. Their theoretical account was based on distributed representations of words. When a person encounters a word, its different components are represented as different vectors in memory. For example, given the prime *SOIL*, the self-generated word has the “meaning” vector *dirt*. However, the presented target has the “perceptual” vector *dart*. Because these are simultaneously experienced, the meaning vector *dirt* is superimposed on the perceptual vector *dart*, and both are convolved with the same “context” vector. During recognition, people false alarm to *dirt* because the probe vector closely matches the blended memory vector (see also Dennis & Humphreys, 2001).

In this study, the test-similar conditions were most important. People made judgments to presented homophones (e.g., *PAWS*) while maintaining memory loads of words that resembled the homophones’ alternate spellings (e.g., *CAUSE*). Words in the memory loads are analogous to the self-generated words in the Humphreys et al. (2001) study. Because memory loads were maintained during the study task, each memory word and its mated homophone would be experienced in the same time frame. Theoretically, when the study homophone (e.g., *PAWS*) is shown, its semantic, phonological, and correct orthographic dimensions are activated (Van Orden & Goldinger, 1994). However, its nonpresented, alternate spelling (e.g., *PAUSE*) is partially activated as well (Stone et al., 1997). In the test-similar condition, the memory load words include orthographic information similar to those alternate spellings (e.g., *CAUSE*). All these dimensions will be associated with the same context vector because they are active at the same time. During recognition, the probe vector (*PAUSE*) is compared with the memory vector (which contains aspects of both *PAWS* and *CAUSE*). In this case, the probe vector will exactly match on the phonological component and will closely match on the orthographic component, increasing false alarms. Simi-

lar to Humphreys et al. (2001), the results indicate that separate events can result in blended episodic memories. In the study-similar condition, the slightly activated alternate spelling (PAUSE) receives no boost from its orthographic vector, making it a less compelling lure.

Another interesting finding was the strong inverse relationship between memory load recall and false alarm rates in the test-similar condition in Experiment 1 (and partially replicated in Experiment 2). Memory load performance presumably relates to WM ability—the more words people can remember during the dual task, the better their WM. The results showed that, as WM decreased, later false alarms increased. This finding has two important implications: First, people with stronger WM seem better able to separate memory load words and study homophones. That is, they appear less susceptible to memory blends, or source errors. Second, although people with weaker WM explicitly recalled fewer memory load words, they experienced a *stronger* impact of those words in the later recognition test. With fewer WM resources, these individuals appear less able to segregate memory load and study words and, thus, create more memory blends (i.e., false alarms) in the recognition test.

To summarize, in this study, we induced specific memory errors by having people maintain memory loads while making semantic judgments about homophones. When memory loads contained words that resembled the homophones' alternate spellings, people were more likely to false alarm to the nonpresented homophones in a later recognition test. These false alarms seem to reflect the blending of separate, but related, lexical events.

REFERENCES

- ARNDT, J., & HIRSHMAN, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory & Language*, **39**, 371-391.
- COLLINS, A. M., & LOFTUS, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**, 407-428.
- DEESE, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, **58**, 17-22.
- DENNIS, S., & HUMPHREYS, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, **108**, 452-478.
- DONALDSON, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, **121**, 275-278.
- HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.
- HUMPHREYS, M. S., BURT, J. S., & LAWRENCE, S. (2001). Expecting dirt but saying dirt: The creation of a blend memory. *Psychonomic Bulletin & Review*, **8**, 820-826.
- JARED, D., & SEIDENBERG, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General*, **120**, 358-394.
- KREUZ, R. J. (1987). The subjective familiarity of English homophones. *Memory & Cognition*, **15**, 154-168.
- LUKATELA, G., LUKATELA, K., & TURVEY, M. T. (1993). Further evidence for phonological constraints on visual lexical access: TOWED primes FROG. *Perception & Psychophysics*, **53**, 461-466.
- LUKATELA, G., & TURVEY, M. T. (1994). Visual access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, **123**, 107-128.
- LUO, C. R., JOHNSON, R. A., & GALLO, D. A. (1998). Automatic activation of phonological information in reading: Evidence from the semantic relatedness decision task. *Memory & Cognition*, **26**, 833-843.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, **114**, 159-188.
- PESTA, B. J., MURPHY, M. D., & SANDERS, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 328-338.
- PEXMAN, P. M., LUPKER, S. J., & JARED, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 139-156.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, **9**, 123-127.
- RUECKL, J. G., & MATHEW, S. (1999). Implicit memory for phonological processes in visual stem completion. *Memory & Cognition*, **27**, 1-11.
- SHIFFRIN, R. M., & STEYVERS, M. (1997). A model for recognition memory: REM—retrieving effectively from memories. *Psychonomic Bulletin & Review*, **4**, 145-166.
- SOMMERS, M. S., & LEWIS, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory & Language*, **40**, 83-108.
- STARK, C. E. L., & MCCLELLAND, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 945-972.
- 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.
- 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 perception of printed words. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1269-1291.
- WATSON, J. M., BALOTA, D. A., & ROEDIGER, H. L., III (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory & Language*, **49**, 95-118.
- ZIEGLER, J. C., MONTANT, M., & JACOBS, A. M. (1997). Feedback consistency effect in lexical decision and naming. *Journal of Memory & Language*, **37**, 533-554.

(Manuscript received March 14, 2003;
revision accepted for publication September 15, 2003.)