Available online at www.sciencedirect.com
SCIENCE DIRECT®
ACADEMIC PRESS
Journal of Memory and Language 49 (2003) 95–118
Journal of Memory and Language
www.elsevier.com/locate/jml

Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks

Jason M. Watson,* David A. Balota, and Henry L. Roediger III

Department of Psychology, Washington University in St. Louis, One Brookings Hall, St. Louis, MO 63130-4899, USA

Received 30 September 2002

Abstract

The present experiments assessed false memories for critical items (e.g., *dog*) following the presentation of semantic associates (e.g., *hound*, *puppy*), phonological associates (e.g., *log*, *dot*), or hybrid lists of semantic and phonological associates (e.g., *hound*, *dot*). Experiment 1 indicated that the addition of only three phonological associates to a list of 10 semantic associates doubled the recall of non-presented critical items. Experiment 2 indicated that embedding 18 semantic associates and 18 phonological associates within a single list led to greater false recall than the sum of the false recall effects produced by lists containing 18 semantic associates or lists containing 18 phonological associates. Experiment 3 explored this over-additivity and indicated that lists of 8 semantic associates and 8 phonological associates produced greater false recall and false recognition than the average produced in lists containing 16 semantic or 16 phonological associates. These studies provide converging evidence that lists of semantic and phonological associates produce over-additive false recall and false recognition of non-presented critical items relative to pure semantic or pure phonological lists. An activation-monitoring framework is presented that provides an account of the increased false memories elicited by hybrid lists of semantic and phonological associates. Experiment 3 also showed that the experience of remembering during recognition was driven more by semantic factors, whereas the experience of knowing was driven more by phonological factors.

Keywords: False memories; DRM paradigm; Semantic processing; Phonological processing; Source monitoring; Associative networks

There has been considerable recent interest in the nature of falsely remembering information that was never directly presented to an individual. This interest has been nurtured in part by Roediger and McDermott (1995) who adapted an experimental technique originally developed by Deese (1959), hereafter referred to as the DRM paradigm. Roediger and McDermott presented participants with lists of 15 words that were the strongest associates to a missing word or critical item in word association norms (Russell & Jenkins, 1954). For example, participants might hear the following list of words, all of which are associated to the non-presented critical item *sleep*: *bed*, *rest*, *awake*, *tired*, *dream*, *wake*, *snooze*, *blanket*, *doze*, *slumber*, *snore*, *nap*, *peace*, *yawn*, *drowsy*. Immediately following the presentation of the study list, participants recalled as many of the list words as possible in any order (i.e., free recall), and were warned against guessing. Despite this warning, participants recalled the non-presented critical items with about the same probability as items that appeared in the middle of the study list. Roediger and McDermott also found high levels of false alarms for the non-presented critical lures in recognition performance. Interestingly, when asked to make "remember–know" metamemory judgments (see Tulving, 1985),

* Corresponding author. Fax: 1-314-935-7588.
E-mail address: jmwatson@artsci.wustl.edu (J.M. Watson).

0749-596X/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0749-596X(03)00019-6

participants claimed to "remember" the actual occurrence of the non-presented critical lures at about the same level as words that were actually presented. These remarkably high levels of false recall and false recognition in the DRM paradigm have been widely replicated and extended (see Roediger, McDermott, & Robinson, 1998, for a partial review).

Extending the DRM paradigm: Phonological false memories

Several researchers extended the scope of the DRM paradigm by constructing lists of phonological associates that converge upon non-presented critical items (e.g., Chan, McDermott, Watson, & Gallo, 2003; McDermott & Watson, 2001; Schacter, Verfaellie, & Anes, 1997; Sommers & Lewis, 1999). For simplicity, in the present study, we use the terms "semantic" or "phonological" to describe the composition of a list of associates. However, it is important to note that semantic lists may include lexical and/or semantic associates (e.g., *lion* and *tiger*) and phonological lists may include orthographic and/or phonological associates (e.g., *log* and *dog*). The notion is that just as there are networks of semantic associates that capture the meaning of words, there are networks of orthographic associates that capture the spelling of words and networks of phonological associates that capture the pronunciation of words (e.g., Collins & Loftus, 1975; Luce & Pisoni, 1998). For example, Sommers and Lewis (1999, Experiment 1) presented lists of words for study, all of which were associated to the non-presented critical item *but*: *buff*, *put*, *bet*, *boot*, *buck*, *hut*, *bud*, *bus*, *buzz*, *bug*, *boat*, *cut*, *bat*, *bit*, *bought*. The phonological associates were generated by adding, deleting, or substituting phonemes from critical items. Consistent with the semantic false memory findings in the DRM paradigm, robust levels of false recall and false recognition have been obtained with lists of phonological associates. Sommers and Lewis reported false recall and false recognition of critical items to be .54 and .67, respectively, following auditory presentation of phonological lists. However, unlike semantic false recognition, which is characterized by a high proportion of "remember" relative to "know" responses (cf. Roediger & McDermott, 1995), early evidence suggests that phonological false recognition is characterized by a roughly equivalent proportion of "remember" and "know" responses (see Schacter et al., 1997).

Creating false memories with pure and hybrid lists of associates

In the present paper, we investigated the influence of combining both semantic and phonological associates in a single list, creating false memories with "hybrid" lists of associates (cf. Watson, Balota, & Sergent-Marshall, 2001). Although we will delay theoretical discussion until the results have been provided, a priori, it is clear that hybrid lists afford a unique opportunity to investigate the roles of different types of associative information in producing false memories. As noted, researchers have typically either looked at pure lists of semantic associates or pure lists of phonological associates, with the vast majority of studies in this area addressing the powerful effects of pure lists of semantic associates. Although semantic lists provide considerable support for the role of conceptual level representations in producing false memories, these lists provide relatively little support for the orthographic and phonological codes that typically accompany the presentation of a word. In fact, McDermott (1997) found relatively small (compared to presented items), but reliable, priming effects in a word fragment completion task for non-presented critical items. Hence, a critical source of information (the perceptual event at encoding) is relatively impoverished for the non-presented critical item in semantic lists. Consistent with this notion, Schacter, Israel, and Racine (1999) suggested that the relative distinctiveness of studied items and non-presented critical items could be used to reduce false memories in the DRM paradigm. For example, Norman and Schacter (1997) found that both young and older adults recalled more sensory information for studied items than non-presented critical items. Within a source monitoring framework, subjects could use sensory information to discriminate those items that were presented from those items that were not presented, even though there is considerable conceptual information available to support the retrieval of non-presented critical items. However, when there is strong support for both conceptual information and perceptual information as in the hybrid lists, subjects may no longer be able to use orthographic and phonological information to help make the discrimination between real and false memories.

In contrast to pure semantic lists, consider pure phonological lists. Here, there is considerable activation for surface-level, perceptual information; however, there is little support for conceptual information. Because conceptual processes are more likely to drive long-term memory performance (e.g., Craik & Lockhart, 1972), one might expect the influence of perceptual processes to be relatively short-lived. Hence, again, the intriguing possibility is that when combining the two sources of information (i.e., more perceptually driven orthographic/phonological information with more conceptually driven lexical/semantic information), one might expect increased false memories, because the missing piece of information (i.e., either perceptual codes for pure semantic lists or conceptual codes for pure phonological lists) is now provided during encoding. Therefore, we predict that hybrid lists will produce greater levels of false recall and false recognition than either pure lists of semantic associates or pure lists of phonological associates.

Experiment 1

The goal of Experiment 1 was to determine the contribution of additional phonological associates to false memories in a list that primarily included semantic associates. This study was motivated, in part, by Robinson and Roediger (1997) who parametrically explored the influence of the addition of associates with pure semantic lists. In their study, there was a clear increase in false recall as one added more semantic associates. For example, in their second experiment, they found false recall rates of .15, .20, .25, and .28 for lists with 6, 9, 12, and 15 semantic associates, respectively. In the present study, we presented one set of participants with lists of 10 semantic associates and added 1, 2, or 3 phonological associates (while also holding list length constant). For comparison sake, we also presented a second set of participants with lists of 10 semantic associates and added 1, 2, or 3 semantic associates (while also holding list length constant). In the Robinson and Roediger (1997) study, the addition of three semantic associates increased false recall by approximately .03 to .05 (including a .05 increase from 9 to 12 associates). As we will see, consistent with the predictions outlined above in the Introduction and consistent with the results of the Robinson and Roediger study, the presence of phonological information clearly produced a much larger effect than the influence of the additional semantic associates.

Method

Participants

There were 66 participants in Experiment 1. They were obtained from the Washington University undergraduate population and volunteered, received course credit, or were paid for their participation. Ten of these 66 participants were removed from the analyses on the basis of a post-experiment question-and-answer session which indicated that (1) the participants had been in a false memory experiment prior to the current study and/or (2) the participants had determined the nature of the experimental manipulation (i.e., inducing false memories with pure and hybrid lists of semantic and phonological associates). Thus, the analyses in Experiment 1 were based on a total of 56 participants, which still permitted the complete counterbalancing of materials into the relevant factors, as described below.

Materials

The present study used twenty 15-item lists of words, with each list designed to elicit a false memory for a particular non-presented critical item. To construct the lists, we first referred to the semantic false memory norms of Stadler, Roediger, and McDermott (1999). Twenty critical items were selected from these norms. Each critical item had a large pool of semantic associates based on the Stadler et al. (1999) norms and a large pool of phonological associates that could be generated from the critical item itself, as described below. In some cases, it was necessary to change the critical item for a list of semantic associates to an alternative, thematically similar critical item in order to obtain a set of rhyming associates. For example, the critical item *doctor* from the Stadler et al. norms has few, if any, rhyming neighbors, so the critical item *sick*, which has several rhyming neighbors, including *kick*, *pick*, etc., was used in its place. Ten semantic associates and three rhyming associates were selected from the pool of associates for each of the 20 critical items. In most cases, the semantic associates were the first 10 associates of a critical item as listed in the Stadler et al. norms. If a thematically similar critical item was used in the place of a critical item (e.g., substituting *sick* for *doctor*, as described above), in most cases, the semantic associates were still the first ten associates of the replaced critical item (i.e., *doctor*) as listed in the Stadler et al. norms. In all cases, the phonological associates were both orthographically and phonologically related to their corresponding critical item. The rhyming associates were created by adding, deleting, or substituting one phoneme or more from the initial positions of critical items (e.g., *sick* → *pick*; *trash* → *rash*; *chair* → *pair*; *king* → *bring*). For comparison sake with the rhymes, an additional three semantic associates were selected for each critical item, thereby providing a conceptual replication of the design of the Robinson and Roediger (1997) study.[1]

For counterbalancing purposes, for half of the participants, the 20 lists were divided into four sets of five lists corresponding to four types of associative lists. As shown in Table 1, the four types of lists included (1) semantic lists, *S*, (2) semantic lists with one rhyming associate, *SR1*, (3) semantic lists with two rhyming associates, *SR2*, and (4) semantic lists with three rhyming associates, *SR3*. For the remaining half of the participants, the 20 lists were again divided into four sets of five lists corresponding to four types of associative lists. However, for these participants, the four types of lists were slightly different than those described above and included (1) semantic lists, *S*, (2) semantic lists with one

[1] We thank an anonymous reviewer for suggesting this control condition in Experiment 1.

Table 1
Sample lists of studied items used for the critical item *Sleep* in Experiment 1

Serial position	*List type*							
		S		*SR1*		*SR2*		*SR3*
1		bed		bed		bed		bed
2		rest		rest		rest		rest
3		awake		awake		awake		awake
4		tired		tired		tired		tired
5		dream		dream		dream		dream
6		scrub		scrub		scrub		scrub
7	U	***file***	U	file	R	***weep***	R	***weep***
8		wake		wake		wake		wake
9	U	***load***	R	***keep***	U	load	R	***keep***
10		snooze		snooze		snooze		snooze
11	U	***honor***	U	honor	R	***steep***	R	***steep***
12		dazzle		dazzle		dazzle		dazzle
13		blanket		blanket		blanket		blanket
14		doze		doze		doze		doze
15		slumber		slumber		slumber		slumber
		S		*SS1*		*SS2*		*SS3*
1		bed		bed		bed		bed
2		rest		rest		rest		rest
3		awake		awake		awake		awake
4		tired		tired		tired		tired
5		dream		dream		dream		dream
6		scrub		scrub		scrub		scrub
7	U	***file***	U	file	S	***drowsy***	S	***drowsy***
8		wake		wake		wake		wake
9	U	***load***	S	***snore***	U	load	S	***snore***
10		snooze		snooze		snooze		snooze
11	U	***honor***	U	honor	S	***pillow***	S	***pillow***
12		dazzle		dazzle		dazzle		dazzle
13		blanket		blanket		blanket		blanket
14		doze		doze		doze		doze
15		slumber		slumber		slumber		slumber

Note: The to-be-replaced unrelated words from serial positions 7, 9, and 11 and their substituted rhyming and semantic associates to the critical item are denoted in bold, italic for each list type. U, unrelated; R, rhyme; S, semantic.

additional semantic associate, *SS1*, (3) semantic lists with two additional semantic associates, *SS2*, and (4) semantic lists with three additional semantic associates, *SS3*. In a semantic list, there were 10 semantic associates to a non-presented critical item and five words that were unrelated to the critical item, yielding a 15-item list. The 10 semantic associates occupied serial positions 1–5, 8, 10, and 13–15 with the five unrelated words filling the remaining serial positions (i.e., 6–7, 9, and 11–12). For the remaining six types of associative lists, the 10 semantic associates and the two unrelated words from serial positions 6 and 12 were used with different combinations of unrelated words and semantic or phonological associates of the non-presented critical items. That is, as shown in Table 1, for *SR1*, *SR2*, and *SR3* lists, the three unrelated words in the 7th, 9th, and 11th serial positions were replaced by rhyming associates to the critical items. For example, in the *SR1* list for the critical item *sleep*, the unrelated word in the 9th serial position of the *S* list (i.e., *load*) was replaced by a rhyming associate (i.e., *keep*). In contrast, as shown in Table 1, for *SS1*, *SS2*, and *SS3* lists, the three unrelated words in the 7th, 9th, and 11th serial positions were replaced by semantic associates to the critical items. For example, in the *SS1* list for the critical item *sleep*, rather than being replaced by a rhyming associate, the unrelated word in the 9th serial position of the *S* list (i.e., *load*) was replaced by a semantic associate (i.e., *snore*).

Although the rhyming associates used in this study were selected randomly from the pool of possible phonological associates for each critical item, the three additional semantic associates were selected with respect to their backward associative strength (BAS) to their corresponding critical items (Nelson, McEvoy, & Schreiber, 1999; Roediger, Balota, & Watson, 2001). Specifically, the mean BAS between the 10 semantic associates used in *S* lists and the corresponding critical item across all 20 lists was .20. Furthermore, the mean BAS between the semantic associates substituted into serial positions 7, 9, and 11 and the corresponding critical item across all 20 lists was .28, .20, and .12, respectively. As shown in Table 1, because the average placement of the semantic and rhyming associates was held constant at the 9th serial position across the different types of lists, the mean BAS between the additional semantic associates in *SS1*, *SS2*, and *SS3* lists and the corresponding critical item across all 20 lists was also .20. In this way, in Experiment 1, independent false memory estimates were obtained for the same critical item following three different types of pure lists of semantic associates (i.e., *SS1*, *SS2*, and *SS3*) and three different types of hybrid lists of semantic and phonological associates (i.e., *SR1*, *SR2*, and *SR3*) while holding list length constant at 15 items across all conditions.[2]

Procedure

Participants were tested individually and were given a total of twenty 15-item lists of words. Half of the participants received five lists in each of the following four conditions: *S*, *SR1*, *SR2*, and *SR3*, whereas the remaining half of the participants received five lists in each of the following four conditions: *S*, *SS1*, *SS2*, and *SS3*. All 20 lists were counterbalanced across the four different conditions within each set of participants. However, the 20 lists were presented in the same pseudo-random order for every participant, with the restriction that successive trials could not have lists of the same condition. At the beginning of the experiment, participants were told that they would be given lists of words on a computer screen, and that after each list, they would be given 90 s to recall the words from the most recently presented list. They were then given a recall packet that consisted of instructions and 20 sheets of paper on which to write their responses. Each piece of paper was labeled to correspond to one of the 20 experimental lists (e.g., "List 1").

On each trial, the following sequence of events occurred: (a) participants were instructed to press the ENTER key to begin presentation of a study list, upon which the screen was blanked; (b) a 250 ms blank interval; (c) a study word appeared in uppercase letters in the middle of the computer screen and remained for 1500 ms; (d) the screen was blanked, and steps b through d repeated for the remaining words on a list; (e) after all 15 words in a list had been presented, there was a 250 ms delay followed by an instruction to begin recall, with the computer providing a 90 s countdown for the participant; (f) after the 90 s recall period was complete, a 1 s tone sounded followed by an instruction to stop recall, and there was a 3 s inter-trial interval; (g) the screen was blanked, and steps a through g repeated for the remaining lists of words.

[2] The entire set of stimuli used in Experiment 1 is available upon request.

Results

Fig. 1 displays the false recall probabilities as a function of increasing numbers of associates (0, 1, 2, or 3) and list type (semantic vs hybrid) in Experiment 1. Veridical recall is not shown in Fig. 1 because it did not change as a function of the number of additional associates in either pure semantic (.60) or hybrid lists (.65) after collapsing across all serial positions for studied items. In contrast, as shown in Fig. 1, false recall increased dramatically with increasing numbers of rhyming associates in hybrid lists of semantic and phonological associates, whereas false recall increased slightly with increasing numbers of semantic associates in pure semantic lists. Specifically, false recall increased from a base rate of .24 in pure semantic or *S* lists, to a rate of .34 in *SR1* lists, to a rate of .41 in *SR2* lists, to a rate of .49 in *SR3* lists which contained a mixture of 10 semantic and three rhyming associates to a non-presented critical item. However, false recall increased from a base rate of .21 in pure semantic or *S* lists, to a rate of .24 in *SS1* lists, to a rate of .26 in *SS2* lists before decreasing to a rate of .23 in *SS3* lists which contained 13 semantic associates to a non-presented critical item.

These observations were supported by a 2 (List Type: Pure Semantic vs Hybrid Lists) $\times$ 4 (Number of Associates) $\times$ 2 (Item Type: Veridical vs False) ANOVA, which yielded main effects of the number of associates,

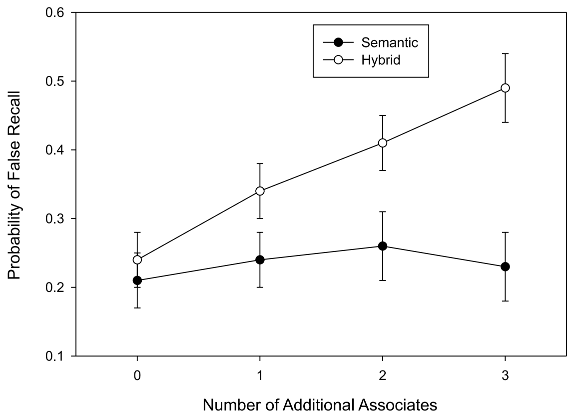

Fig. 1. Mean probability of false recall as a function of increasing numbers of associates and list type in Experiment 1. Error bars represent the standard error of the mean.

$F(3, 162) = 7.25$, $MSe = .018$, $p < .001$, and item type, $F(1, 54) = 147.86$, $MSe = .079$, $p < .001$. The effect of list type was marginally significant, $F(1, 54) = 3.94$, $MSe = .046$, $p < .06$. There were reliable two-way interactions of the number of associates and item type, $F(3, 162) = 3.06$, $MSe = .025$, $p < .05$; number of associates and list type, $F(3, 162) = 2.93$, $MSe = .018$, $p < .05$; and list type and item type, $F(1, 54) = 13.88$, $MSe = .079$, $p < .001$. Most importantly, there was a reliable three-way interaction of list type, number of associates, and item type, $F(3, 162) = 3.29$, $MSe = .025$, $p < .05$. An additional 2 (List Type) × 4 (Number of Associates) ANOVA for non-presented, non-critical item intrusions revealed no significant effects (all Fs < 1). The intrusion of other non-presented, non-critical items was very low in Experiment 1 (≅.32 items/list).

Discussion

The results from Experiment 1 are quite clear. Specifically, there was a dramatic increase in false recall by simply adding 1, 2, or 3 phonologically related words. In fact, false recall of critical items doubled from a baseline probability of .24 with pure semantic lists to a probability of .49 with hybrid lists that consisted of ten semantic associates plus three phonological associates. In contrast, there was a slight increase in false recall with three additional semantic associates. Specifically, false recall of critical items increased from a baseline probability of about .21 with 10 semantic associates to a probability of about .23 with 13 total semantic associates. It is interesting to note that Robinson and Roediger (1997) only observed a .05 increase in false recall when the total number of semantic associates increased by three semantic associates from 9 to 12 items. Although there are problems with cross-study comparisons, it is also noteworthy that the 10 semantic associates in the present study produced comparable baseline false recall performance (.23) to the interpolated estimate (.22) from the Robinson and Roediger study. In this light, it is quite intriguing that the addition of *six* semantic associates in the Robinson and Roediger study (i.e., from 9 to 15) only produced an increase in false recall of .08, whereas, a *single* phonological associate in the present study produced an increase of .10 (and, as noted, three phonological associates produced an increase of .25). Thus, the results of Experiment 1 clearly (1) replicated the false memory results of Robinson and Roediger, and more importantly, (2) extended their findings by demonstrating an exaggerated influence in false recall by adding a few phonological associates (instead of semantic associates) to a list of semantic associates.

Experiment 2

Although the results from Experiment 1 suggest that hybrid lists of associates produce exaggerated false recall effects, in Experiment 2, we will more directly investigate the nature of false memories elicited by hybrid lists by determining if the effect of mixing semantic and phonological associates is additive, under-additive, or over-additive (see Balota & Paul, 1996, for a discussion of additivity issues). In order to address this issue, participants studied 72-item lists of words that contained subsets of associates that were semantically (e.g., *hound*, *puppy*), phonologically (e.g., *log*, *dot*), or both semantically and phonologically (e.g., *hound*, *dot*) related to a non-presented critical word (e.g., *dog*). These 72-item lists included (a) pure sets of 18 semantic associates, (b) pure sets of 18 phonological associates, or (c) hybrid sets of 36 semantic and phonological associates. In a simple additive model, the probability of false recall would only be influenced by the amount of activation for the critical item that had accumulated across independent semantic and phonological associative networks. In contrast, because a hybrid list converges upon and uniquely constrains the critical item (via the arbitrary mapping of semantic and phonological information), one might predict an over-additive influence of the semantic and phonological associates. If this is the case, then false recall with the hybrid lists will be *greater* than the sum of the false recall effects observed in the pure semantic and pure phonological lists.

Method

Participants

There were 52 participants in Experiment 2. They were obtained from the Washington University undergraduate population and received course credit for their participation. Four participants were removed from the analyses on the basis of a post-experiment question-and-answer session which indicated that (1) the participants had been in a false memory experiment prior to the current study and/or (2) the participants had determined the nature of the experimental manipulation (i.e., inducing false memories with pure and hybrid lists of semantic and phonological associates). Thus, the analyses in Experiment 2 were based on a total of 48 participants, which still permitted the complete counterbalancing of materials into the relevant factors, as described below.

Materials

The present study used six 72-item lists of words, with each list designed to elicit a false memory for a particular non-presented critical item. Based on the Nelson et al. (1999) norms and two different rhyming dictionaries (Carnegie Mellon Semantic-Rhyming Dictionary, on-line; Redfield, 1986), each of these critical items has a large pool of semantic and phonological associates, respectively. Eighteen semantic and 18 phonological associates were selected from this pool of associates for each of the six critical items. In order to allow a broader selection of phonological associates in Experiment 2, based on prior work by Sommers and Lewis (1999), phonologically related words were created by adding, substituting, or deleting phonemes from all positions in the critical item (e.g., *log*, *dot*, *dug* → *dog*).

For counterbalancing purposes, the six lists were divided into three sets of two lists corresponding to the three types of associative lists used in the present study. The three types of associative lists included (1) pure

Table 2
Sample lists of studied items used for the critical item *Dog* in Experiment 2

Serial position	*List type*					
	S		*P*		*H*	
1–6	U	audit,...	U	audit,...	U	audit,...
7	S	***vet***	U	melt	S	***vet***
8	U	ask	P	***log***	P	***log***
9	S	***puppy***	U	ounce	S	***puppy***
10	U	egg	P	***dug***	P	***dug***
11	S	***mutt***	U	wasp	S	***mutt***
12	U	blur	P	***bog***	P	***bog***
13	S	***bark***	U	tour	S	***bark***
14	U	reef	P	***daub***	P	***daub***
15	S	***flea***	U	rink	S	***flea***
16	U	roam	P	***dock***	P	***dock***
17	S	***tail***	U	camp	S	***tail***
18	U	end	P	***hog***	P	***hog***
19–30	U	coin,...	U	coin,...	U	coin,...
31	S	***hound***	U	wedge	S	***hound***
32	U	limp	P	***jog***	P	***jog***
33	S	***animal***	U	speech	S	***animal***
34	U	noun	P	***doll***	P	***doll***
35	S	***pooch***	U	gravy	S	***pooch***
36	U	tease	P	***dodge***	P	***dodge***
37	S	***cat***	U	loop	S	***cat***
38	U	rip	P	***dag***	P	***dag***
39	S	***beware***	U	hurdle	S	***beware***
40	U	moss	P	***frog***	P	***frog***
41	S	***bite***	U	wrap	S	***bite***
42	U	raft	P	***dig***	P	***dig***
43–54	U	notice,...	U	notice,...	U	notice,...
55	S	***collar***	U	silver	S	***collar***
56	U	oven	P	***dawn***	P	***dawn***
57	S	***Lassie***	U	amplify	S	***Lassie***
58	U	age	P	***cog***	P	***cog***
59	S	***pet***	U	hem	S	***pet***
60	U	gym	P	***dot***	P	***dot***
61	S	***poodle***	U	insert	S	***poodle***
62	U	hike	P	***doff***	P	***doff***
63	S	***bone***	U	disk	S	***bone***
64	U	tug	P	***pog***	P	***pog***
65	S	***paw***	U	lift	S	***paw***
66	U	bib	P	***fog***	P	***fog***
67–72	U	council,...	U	council,...	U	council,...

Note: The semantic and phonological associates to the critical item from serial positions 7 to 18, 31 to 42, and 55 to 66 are denoted in bold, italic for each list type. U, unrelated; S, semantic; P, phonological.

semantic lists, *S*, (2) pure phonological lists, *P*, and (3) hybrid lists, *H*. The three types of associative lists are illustrated in Table 2 for the critical item *dog*. In a pure semantic list, there were 18 semantic associates to a non-presented critical item and 54 words that were unrelated to the critical item. The 18 semantic associates occurred in three 6-item clusters, occupying the odd-numbered slots in the range of serial positions from 7 to 18, 31 to 42, and 55 to 66 (i.e., serial positions 7, 9, 11, etc.). Eighteen of the unrelated items in pure semantic lists were matched on length and log frequency to the phonological associates of the corresponding critical item. The 18 matched items occurred in three 6-item clusters, occupying the even-numbered slots in the range of serial positions from 7 to 18, 31 to 42, and 55 to 66 (i.e., serial positions filled by phonological associates in pure phonological lists).

Turning to the phonological lists, the 18 phonological associates occurred in three 6-item clusters, occupying the even-numbered slots in the range of serial positions from 7 to 18, 31 to 42, and 55 to 66 (i.e., serial positions 8, 10, 12, etc.). Eighteen of the unrelated items in pure phonological lists were matched on length and log frequency to the semantic associates of the corresponding critical item. The 18 matched items occurred in three 6-item clusters, occupying the odd-numbered slots in the range of serial positions from 7 to 18, 31 to 42, and 55 to 66 (i.e., serial positions filled by semantic associates in pure semantic lists). For both pure semantic and pure phonological lists, thirty-six unrelated words filled the remaining serial positions. These unrelated words were further subdivided into 6- and 12-item clusters located in the range of serial positions from 1 to 6, 19 to 30, 43 to 54, and 67 to 72.

For the hybrid lists, the same semantic and phonological associates from the pure lists occurred in three 6-item clusters, occupying the same odd-numbered slots and even-numbered slots, respectively, in the range of serial positions from 7 to 18, 31 to 42, and 55 to 66. In addition, a hybrid list filled the 36 serial positions from 1 to 6, 19 to 30, 43 to 54, and 67 to 72 with the same 6- and 12-item clusters of unrelated words from the pure lists of associates. In this way, unlike pure lists that had 18 semantic or 18 phonological associates to a non-presented critical item, hybrid lists had 36 associates that converged on a non-presented critical item across both semantic and phonological domains.[3]

Procedure

Participants were tested either individually or in groups that ranged in size from 2 to 4 people. They were given six 72-item lists of words, with two lists in each of the three different conditions (i.e., *S*, *P*, and *H*). The lists were counterbalanced across the three different conditions across participants. The six lists were presented in a random order for each participant. At the beginning of the experiment, participants were told that they would receive six lists of 72 words presented on a computer screen, and that after each list, they would be given 6 min to recall words from the most recently presented list. They were then given a recall packet that consisted of instructions and sheets of paper on which to write their responses. Each piece of paper was labeled to correspond to one of the six experimental lists (e.g., "List 1").

On each trial, the following sequence of events occurred: (a) participants were instructed to press the ENTER key to begin presentation of a study list, upon which the screen was blanked; (b) a 250 ms blank interval; (c) a study word appeared in uppercase letters in the middle of the computer screen and remained for 1000 ms; (d) the screen was blanked, and steps b through d repeated for the remaining words on a list; (e) after all 72 words in a list had been presented, there was a 250 ms delay followed by an instruction to begin recall, with the computer providing a 6 min countdown for the participant; (f) after the 6 min recall period was complete, a 1 s tone sounded followed by an instruction to stop recall, and there was a 3 s inter-trial interval; (g) the screen was blanked, and steps a through g repeated for the remaining lists of words.

Results

Fig. 2 displays the veridical and false recall probabilities by item type for each of the three list types used in Experiment 2: semantic (*S*), phonological (*P*), and hybrid (*H*). There are four points to note in Fig. 2. First, veridical recall for semantic associates of the critical item (.40) was higher than veridical recall for phonological associates (.24).[4] Second, veridical recall for both types of associates was higher than veridical recall for words not related to the critical item, including (a) the length- and frequency-matched items to the semantic and phonological associates, and (b) the set of unrelated

[3] The entire set of stimuli used in Experiment 2 is available upon request. If an item was not present in the Kucera and Francis (1967) database, it was assigned a raw frequency of 1 when calculating log frequency.

[4] This pattern was also true for hybrid lists of associates, which is an average of the veridical recall probabilities for semantic associates (.36) and phonological associates (.20), respectively.

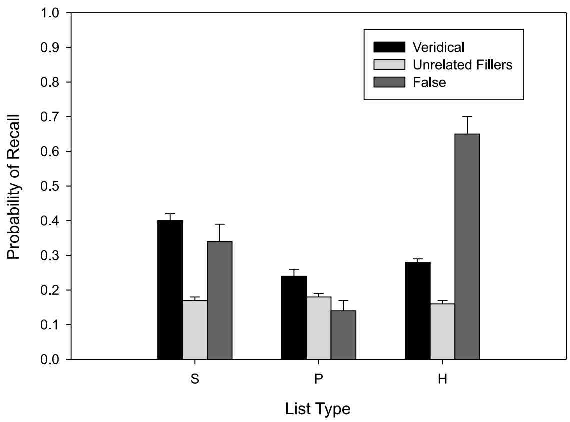

Fig. 2. Mean probability of veridical recall and mean probability of false recall as a function of list type in Experiment 2. Error bars represent the standard error of the mean.

filler items.[5] Third, although veridical recall of associates and unrelated words was fairly stable across pure and hybrid lists, false recall changed dramatically as a function of list type. Specifically, false recall of the critical item was greater with hybrid lists (.65) than with pure semantic lists (.34), which was, in turn, greater than false recall with pure phonological lists (.14). Fourth, and most importantly, false recall with hybrid lists (.65) appeared to be over-additive, because the observed probability was approximately .17 larger than the predicted additive probability from the pure lists (i.e., $.34 + .14 = .48$). Consistent with these observations, the results of a 3 (List Type) $\times$ 3 (Item Type: Associates vs Unrelated Fillers vs Critical Items) ANOVA indicated that there were main effects of list type and item type, $F(2,94) = 30.87$, $MSe = .037$, $p < .001$, and $F(2,94) = 31.79$, $MSe = .049$, $p < .001$, respectively. More importantly, the interaction of list type and item type was also significant, $F(4,188) = 33.46$, $MSe = .035$, $p < .001$.

A series of ANOVAs were conducted on false recall measures to determine whether false recall was greater with hybrid lists than with pure lists. The first ANOVA simply confirmed that there was a highly reliable effect of list type on false recall, $F(2,94) = 32.38$, $MSe = .098$, $p < .001$. The second ANOVA indicated that the intrusion of other non-presented, non-critical items did not vary as a function of list type, $F(2,94) = 1.37$, $MSe = 1.92$, $p = .26$. Finally, an ANOVA on the predicted false recall versus the observed false recall in hybrid lists for each subject indicated that the observed false recall following hybrid lists of associates (.65) was larger than that predicted by the sum of the false recall effects in pure phonological and pure semantic lists (.48), $F(1,47) = 5.61$, $MSe = .119$, $p < .05$.

Discussion

The results of Experiment 2 replicated and extended the results of Experiment 1. Specifically, in both studies, false recall of a non-presented critical item dramatically increased with hybrid lists of semantic and phonological associates compared to pure lists of semantic associates. Furthermore, as shown in Experiment 2, false recall with hybrid lists was over-additive with respect to the predicted sum of false recall observed with pure semantic and pure phonological lists. In fact, to our knowledge, the .65 recall of critical items that was observed with hybrid lists of associates is the highest level of false recall ever reported in the DRM false memory literature. This robust, over-additive false recall may be due, in part, to breakdowns in source monitoring with hybrid lists of associates. For example, as additional semantic and phonological associates are presented in the context of a 72-item list, it may be increasingly difficult for participants to determine whether the critical item was presented or whether items that looked, sounded, or meant something similar to the critical item were presented (especially with respect to studied items presented at the beginning of the 72-item list).

Experiment 3

At this point, it is noteworthy that in the DRM false memory literature, participants typically study much shorter lists (e.g., 15-item lists of associates) than the 72-item lists of associates used in Experiment 2. Therefore, participants in Experiment 2 may have had particular difficulty attempting to monitor the source of information and to determine that critical items were activated but not presented. Thus, one goal of Experiment 3 was to attempt to replicate the over-additive pattern observed in Experiment 2 with 16-item lists of semantic and/or phonological associates.

A second goal of Experiment 3 was to test the additivity hypothesis in a different manner. Specifically, in Experiment 2, there were 36 total associates in the

[5] Only veridical recall for the unrelated filler items (i.e., the unrelated words located in the range of serial positions from 1 to 6, 19 to 30, 43 to 54, and 67 to 72) is shown in Fig. 2. However, veridical recall probabilities for the unrelated words that were length- and frequency-matched to the semantic (.13) and phonological associates (.12) were comparable to the recall of these unrelated filler items (.17). The small difference in recall performance for these two different classes of unrelated stimuli ($\cong$.04) may be due to the influence of the primacy (i.e., 1–6) and recency (i.e., 67–72) serial positions for the unrelated filler items.

hybrid lists, but only 18 total associates in pure semantic or pure phonological lists. Thus, one might argue that the elevated, over-additive false recall observed for hybrid lists could be attributed to the increased total number of associates for hybrid lists relative to pure lists of associates. To address this possibility, in Experiment 3, false recall was again assessed following the three types of lists: pure semantic, pure phonological, and hybrid lists of semantic and phonological associates. However, unlike Experiment 2, list type was not correlated with the total number of associates in Experiment 3. Rather, for all three conditions, the total number of associates was equated by using lists with 16 semantically associated items, 16 phonologically related items, or hybrid lists with 8 semantically associated and 8 phonologically related items. Thus, if there are additive effects of semantic and phonological associates, then false recall with hybrid lists will *equal* the average of the false recall effects observed with the pure semantic and pure phonological lists, because there are an equivalent number of total associates across the three different conditions. In contrast, if there are over-additive effects of semantic and phonological associates, then false recall with hybrid lists will be *greater* than the average of the false recall effects observed in pure lists of associates.

Finally, Experiment 3 also included veridical and false recognition estimates for pure and hybrid lists of associates, including "remember–know" judgments. This allowed for further tests of the additivity of false memories in recognition (e.g., false recognition with and without prior recall). Moreover, as noted in the Introduction, unlike semantic false recognition, which is characterized by a high proportion of "remember" relative to "know" responses (cf. Roediger & McDermott, 1995), early evidence suggests that phonological false recognition is characterized by a roughly equivalent proportion of "remember" and "know" responses (see Schacter et al., 1997).

Method

Participants

There were 36 participants in Experiment 3. They were obtained from the Washington University undergraduate population and received course credit for their participation. Four of these 36 participants were removed from the analyses because they failed to follow the instructions of the study. Specifically, on multiple occasions, each of these participants recalled a list of words, even though the computer had instructed them to solve math problems. Moreover, based on a post-experiment question-and-answer session, one of these four participants also had some knowledge of the experimental manipulation. Thus, the analyses in Experiment 3 were based on a total of 32 participants.

Materials

The present study used thirty-six 16-item lists of words. Based on the Nelson et al. (1999) norms and two different rhyming dictionaries (Carnegie Mellon Semantic-Rhyming Dictionary, on-line; Redfield, 1986), critical items were identified that had a large pool of semantic and phonological associates, respectively. In order to allow a broad selection of phonological associates, as in Experiment 2, phonologically related words were created by adding, substituting, or deleting phonemes from all positions in the critical item (e.g., *log*, *dot*, *dug* $\rightarrow$ *dog*). The entire set of stimuli used in Experiment 3 is given in appendix, including the semantic and phonological associates for each list and its corresponding critical item.

For counterbalancing purposes, the 36 lists were subdivided into three sets of 12 lists corresponding to the three types of associative lists used in the present study. The three types of associative lists included (1) pure semantic lists, *S*, (2) pure phonological lists, *P*, and (3) hybrid lists, *H*. These three sets of 12 lists were further sub-divided into three sets of four lists that represented the test condition that would be received for a particular list. Specifically, for four of the 12 lists in each condition, there was a recall test immediately following the presentation of the associates; however, for another four of the lists, a set of addition, subtraction, and multiplication problems were given in place of an immediate recall test. For the remaining four lists, the associates were not presented, and, thus, were not followed by an immediate recall test or math problems. Finally, for all of the 36 lists, a surprise recognition test was given at the end of the experiment. Therefore, the present study used a 3 (List Type: *S*, *P*, or *H*) $\times$ 3 (Test Type: study and recall, study and math, no study) design, yielding four observations per participant per cell. This design permitted estimates of immediate recall as well as three different estimates of delayed recognition (i.e., preceded by recall, not preceded by recall, and non-studied baseline) following the three types of associative lists. Critical items and the corresponding semantic and phonological associates were counterbalanced across the three different list types and the three different test types across participants.

The three types of associative lists are illustrated in Table 3 for the critical item *dog*. As shown, there were two types of hybrid lists (H_1 and H_2) for each critical item that were created by interleaving 8 semantic and 8 phonological associates from its corresponding pure lists (e.g., *hound*, *dodge*, *bite*, *hog*,...). The two lists were composed either by selecting the odd-numbered serial positions in a pure list or the even-numbered serial

Table 3
Sample lists of studied items used for the critical item *Dog* in Experiment 3

Serial position	*List type*							
	S		*P*		H_1		H_2	
1	S	hound	P	log	S	hound	P	log
2	S	puppy	P	dodge	P	dodge	S	puppy
3	S	bite	P	dug	S	bite	P	dug
4	S	mutt	P	hog	P	hog	S	mutt
5	S	pet	P	bog	S	pet	P	bog
6	S	beware	P	doff	P	doff	S	beware
7	S	bone	P	daub	S	bone	P	daub
8	S	tail	P	cog	P	cog	S	tail
9	S	cat	P	dock	S	cat	P	dock
10	S	animal	P	dawn	P	dawn	S	animal
11	S	paw	P	fog	S	paw	P	fog
12	S	poodle	P	dig	P	dig	S	poodle
13	S	flea	P	doll	S	flea	P	doll
14	S	bark	P	frog	P	frog	S	bark
15	S	Lassie	P	jog	S	Lassie	P	jog
16	S	vet	P	dot	P	dot	S	vet

Note: S, semantic; P, phonological.

positions. For the purpose of the analyses, the two subtypes of hybrid lists were combined together to form a single condition for hybrid lists.

The delayed recognition test consisted of 108 items presented in a single random order. This set included 36 critical items, 36 semantic associates, and 36 phonological associates. For each participant, 24 of the items on the recognition test were old (including 12 semantic and 12 phonological associates that were studied, with half of each set of 12 items followed by recall and the other half followed by math), and 84 of the items on the recognition test were new (including the remaining 24 semantic and 24 phonological associates that were not studied, and all 36 critical items). Critical items were never presented during the study phase of the experiment. The semantic and phonological associates for each critical item were drawn from the 9th serial position from the pure lists of associates shown in appendix; however, depending on the list a given participant received, these items may or may not have been presented during the study phase of the experiment.

Procedure

Participants were tested either individually or in groups that ranged in size from 2 to 15 people. Participants were presented with twenty-four 16-item lists of words, with four lists in each of the six different cells produced by factorially crossing List Type (i.e., *S*, *P*, and *H*) with Test Type (i.e., study and recall, study and math). The 24 lists were presented in a random order for each participant. The remaining twelve 16-item lists of associates were not presented to participants (i.e., no study), with four lists in each of the three different list types. At the beginning of the experiment, participants were told that they would be asked to memorize lists of words that were presented to them on a computer screen. They were then given a recall packet that consisted of instructions and twelve sheets of paper on which to write their responses. Each piece of paper in the recall packet was labeled to correspond to one of the 12 experimental lists (e.g., "List 1") with 16 blank spaces, which corresponded to the total number of words presented on a single list. Participants were also given a math packet that consisted of instructions and sets of math problems. Each piece of paper in the math packet was labeled to correspond to one of the 12 presented lists that was not to-be-recalled (e.g., "Problem Set #1"). Furthermore, each piece of paper contained a different set of 18 total math problems, including 6 addition, 6 subtraction, and 6 multiplication problems.

Participants were told that they would either be asked to remember the list of words that was just shown or to solve math problems. Since participants did not know whether the computer would instruct them to recall words or to solve math problems until after a list was presented, they were encouraged to memorize each list of words to the best of their ability in anticipation of a possible memory test. Participants were told that they would have 60 s after each list to recall the words or to solve math problems. More importantly, participants were told not to guess when trying to remember the words from the lists and to only write down words if they were sure that those words had been presented. Participants were also told that some of the words would be in uppercase letters and that some of the words would be in lowercase letters. However, participants were told that case was unimportant and were instructed to ignore case when trying to remember the words from a list.

At the end of the math packet, there was a surprise 108-item recognition test. This recognition test was administered at the end of the experiment, after the recall/math portion of the experiment was complete. All of the stimuli on the recognition test were presented in uppercase letters, consistent with the uppercase semantic and rhyming associates drawn from the 9th serial position in pure lists. The instructions for the recognition test were as follows:

> "In this portion of the experiment, you will be given a recognition test. Some of the words presented on this recognition test are ***OLD*** items which were presented earlier during the first portion of this experiment. However, some of the recognition test words are ***NEW*** items which were not presented earlier in this experiment. In the margin beside each item, please circle the word ***OLD*** if you judge the item as having been presented earlier during this experiment. Please circle the word ***NEW*** if you judge the item as having not been presented earlier in this experiment.
>
> In addition to the ***OLD/NEW*** judgment, for those items that you judge to be ***OLD***, I would like you to make an additional judgment. Specifically, I would like to know whether you ***REMEMBER*** having seen the word on the computer screen or whether you just ***KNOW*** that it was presented. Please allow me to explain the distinction.
>
> A ***REMEMBER*** response would be made for items which you can remember the ***specific*** context in which it was presented. For example, you might remember the word that came immediately before or after it, or you might remember what the word made you think of when you saw it.
>
> A ***KNOW*** response would be made for items which you cannot recollect any specific aspect of the presentation episode, but that you ***just know*** were on the list. You know that you saw it on the computer screen, but you don't remember anything specific about the event.
>
> For a real life example, suppose you saw me in the grocery store next week. You might recognize me in one of two ways. You could ***REMEMBER*** me as the experimenter you met last week at the Psychology Building, or you could just ***KNOW*** that you had seen me recently without remembering the context.
>
> So, please circle an ***R*** for ***REMEMBER*** or circle a ***K*** for ***KNOW*** next to each item you identify as ***OLD***. Also, the ***REMEMBER/KNOW*** judgment should be made with respect to the study episode, that is, when the words were presented on the computer screen, and not the recall tests. In other words, I am not interested in whether you ***REMEMBER*** or ***KNOW*** that you wrote down an item during the recall test but whether you ***REMEMBER*** or ***KNOW*** that it was given to you in one of the study lists.
>
> Please feel free to ask the experimenter any questions that you may have at this point. If you feel that you understand these instructions, you may begin this portion of the experiment. You may take as long as you like in completing this recognition test."

On each trial of the recall/math portion of the experiment, the following sequence of events occurred: (a) participants pressed the ENTER key to begin presentation of a study list, upon which the screen was blanked; (b) a 500 ms delay; (c) a 300 ms tone; (d) a 32 ms inter-stimulus interval; (e) a study word appeared in either uppercase (odd-numbered serial positions) or lowercase letters (even-numbered serial positions) in the middle of the computer screen and remained for 1000 ms; (f) the screen was blanked, and steps d through f repeated for the remaining words on a list; (g) after all 16 words in a list had been presented, there was a 500 ms delay followed by an instruction to begin recall or math problems, with the computer providing a 1 min countdown for the participant; (h) after the 1 min recall period was complete, a 1 s tone sounded followed by an instruction to stop recall or math problems, and there was a 5 s inter-trial interval; (i) the screen was blanked, and steps a through i repeated for the remaining lists of words.

Results

Recall performance

Fig. 3 displays the probability correct serial recall curves for each of the three list types used in Experiment 3: semantic, phonological, and hybrid. Each of the recall curves corresponded roughly to the typical bow-shaped serial position function. However, there appeared to be some separation in these three recall curves, particularly for semantic lists relative to phonological and hybrid

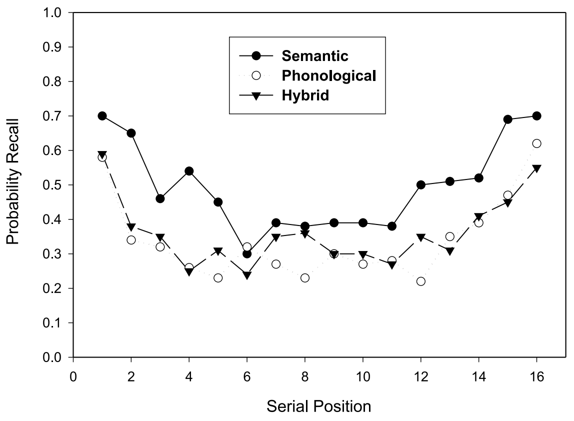

Fig. 3. Mean probability of correct recall as a function of serial position and list type in Experiment 3.

lists. Consistent with these observations, the results of a 3 (List Type) $\times$ 16 (Serial Position) ANOVA indicated that there was a main effect of serial position on recall performance, $F(15, 465) = 17.14$, $MSe = .064$, $p < .001$. There was also a main effect of list type, $F(2, 62) = 79.6$, $MSe = .046$, $p < .001$. The interaction of serial position and list type was also significant, $F(30, 930) = 1.69$, $MSe = .055$, $p < .05$, reflecting the fact that semantic lists produced larger benefits in recall performance over the phonological and hybrid lists at the primacy and recency portions of the curve.

Turning to the analysis of false recall, Fig. 4 displays the veridical and false recall probabilities as a function of list type [semantic (S), phonological (P), and hybrid (H)], collapsing across serial position for studied items. There are three important points to note in Fig. 4. First, as in Experiment 2, lists of semantic associates (.50) were recalled better than lists of phonological associates (.34).

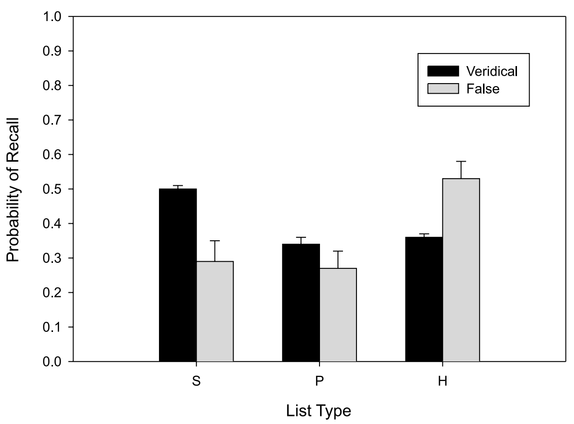

Fig. 4. Mean probability of veridical recall and mean probability of false recall as a function of list type in Experiment 3. Error bars represent the standard error of the mean.

Second, although veridical recall was greater in semantic than in phonological lists, recall of the non-presented critical item remained fairly stable ($\cong$.28). Third, although veridical recall in hybrid lists (.36) decreased relative to pure semantic lists, and was comparable to pure phonological lists, false recall increased dramatically in hybrid lists. Specifically, false recall of the critical item with hybrid lists (.53) was greater than the predicted mean (.28) of the pure semantic lists (.29) or pure phonological lists (.27). Consistent with these observations, the results of a 3 (List Type) $\times$ 2 (Item Type: Veridical vs False) ANOVA indicated that there was a main effect of list type, $F(2, 62) = 7.59$, $MSe = .043$, $p < .001$, but no main effect of item type, $F(1, 31) = 1.04$, $MSe = .058$, $p = .32$, respectively. Most importantly, the interaction of list type and item type was also significant, $F(2, 62) = 13.76$, $MSe = .042$, $p < .001$.

A series of ANOVAs were conducted on false recall measures to determine whether false recall was greater with hybrid lists than with pure lists. First, there was a main effect of list type, $F(2, 62) = 8.25$, $MSe = .082$, $p < .001$. Also, the intrusion of non-presented, non-critical items differed as a function of list type, $F(2, 62) = 4.81$, $MSe = .139$, $p < .05$. Specifically, non-critical item intrusions were highest following phonological lists (.56 items/list), moderate following hybrid lists (.42 items/list), and lowest following semantic lists of associates (.27 items/list). However, for all three list types, the intrusion of non-critical items was relatively low and the pattern of intrusions across list type was inconsistent with the increased probability of false recall following hybrid lists of associates relative to pure lists of semantic or phonological associates. That is, on the basis of differences in sheer intrusion rates alone, one would predict that the highest level of false recall would occur following phonological lists and that the lowest level of false recall would occur following semantic lists. Of course, as shown in Fig. 4, this pattern of results did not occur. The third ANOVA indicated that the recall of the critical item following hybrid lists of associates (.53) was over-additive with respect to the predicted average based on the pure lists (.28), $F(1, 31) = 18.38$, $MSe = .055$, $p < .001$.

Recognition performance

Veridical and false recognition. The top panel of Table 4 displays the probability of "old" responses by item type (veridical vs false) and test type (study and recall, study and math, no study baseline) for each of the three list types used in Experiment 3: semantic, phonological, and hybrid. First, as shown, baseline levels of false recognition were relatively low, and roughly equivalent, for critical items (.17) and unstudied semantic (.14) and

Table 4
Probability of old, remember, and know responses on the recognition test as a function of list type, test type, and item type

	List type								
	Semantic			*Phonological*			*Hybrid*		
	Recall	*Math*	*Base*	*Recall*	*Math*	*Base*	*Recall*	*Math*	*Base*
Old									
Veridical	.69	.63	.14	.57	.38	.15	.54	.41	.15
False	.54	.52	.17	.44	.24	.17	.60	.59	.17
Remember									
Veridical	.47	.39	.03	.19	.14	.03	.35	.20	.03
False	.37	.27	.02	.17	.05	.02	.36	.26	.02
Know									
Veridical	.22	.23	.11	.38	.24	.12	.20	.20	.13
False	.17	.26	.14	.27	.19	.14	.24	.34	.14

Note: Rounding error may prevent the probability of "remember" and "know" responses from summing to equal the exact probability of "old" responses.

phonological associates (.15). Second, collapsing across item type, recognition of associates and critical items appeared to be influenced more by prior recall with phonological lists (.51 for study and recall vs .31 for study and math, respectively) than with either semantic (.61 for study and recall vs .57 for study and math) or hybrid lists (.57 for study and recall vs .50 for study and math). Third, collapsing across recall and math test types, veridical and false recognition varied as a function of list type, with hybrid lists yielding greater recognition of critical items (.60) than studied items (.48), whereas semantic and phonological lists showed the opposite pattern, yielding higher recognition of studied items (.66 and .48, respectively) than critical items (.53 and .34, respectively). Fourth, false recognition of the critical item was greater with hybrid lists (.60) than with pure semantic lists (.53) or pure phonological lists (.34). Thus, as in the false recall data described above, false recognition with hybrid lists appeared to be over-additive, because the observed probability (.60) was approximately .16 larger than the predicted probability (.44) based on the average of the pure semantic and pure phonological lists.

Consistent with the above observations, the results of a 3 (List Type) $\times$ 2 (Test Type: study and recall vs study and math) $\times$ 2 (Item Type) ANOVA indicated that there were main effects of list type, $F(2,62) = 13.53$, $MSe = .084$, $p < .001$, and test type, $F(1,31) = 20.32$, $MSe = .049$, $p < .001$, respectively, but no main effect of item type, $F(1,31) = 2.61$, $MSe = .077$, $p = .12$. More importantly, the two-way interactions of list type and test type, $F(2,62) = 4.72$, $MSe = .045$, $p < .02$, and list type and item type, $F(2,62) = 9.01$, $MSe = .075$, $p < .001$, respectively, were both significant. Neither the two-way interaction of test type and item type, $F(1,31) = 2.31$, $MSe = .031$, $p = .14$, nor the three-way interaction of list type, test type, and item type, $F(2,62) = .69$, $MSe = .059$, $p = .50$, were significant.[6]

A series of ANOVAs were conducted on false recognition measures to determine whether false recognition was greater with hybrid lists than with pure lists of associates. First, there was a main effect of list type, after collapsing across recall and math test types, $F(2,62) = 14.47$, $MSe = .039$, $p < .001$. Additional comparisons indicated that the variability in false recognition of critical items as a function of list type was significant when preceded by math during the study portion of the experiment, $F(2,62) = 19.42$, $MSe = .057$, $p < .001$; however, the effect of list type was only marginally significant when preceded by recall, $F(2,62) = 2.92$, $MSe = .070$, $p < .07$. Turning to the test of additivity, a second ANOVA indicated that the recognition of the critical item following hybrid lists of associates (.60) was over-additive with respect to the predicted average for hybrid lists (.44), $F(1,31) = 21.63$, $MSe = .019$, $p < .001$. It is also worth noting that this over-additive effect was smaller when recognition was preceded by study and recall, because the observed probability (.60) was approximately .11 larger than the predicted average (i.e., $[.54 + .44] \div 2 = .49$), $F(1,31) = 3.92$, $MSe = .050$, $p < .06$, whereas, the over-additive effect was larger when recognition was preceded by study and math, because the observed probability (.59) was approximately .21 larger than the predicted average (i.e., $[.52 + .24] \div 2 = .38$), $F(1,31) = 25.03$, $MSe = .028$, $p < .001$.

"Remember–know" metamemory judgments. Turning to the analysis of metamemory judgments, the middle

[6] Because baseline probabilities of recognition were roughly equivalent as a function of list type and item type, they were ignored for the purposes of the preceding analysis in Experiment 3.

and lower panels of Table 4 display the probability of "remember" and "know" responses, respectively, which sum to equal the probability of "old" responses by item type (veridical vs false) and test type (study and recall, study and math, no study baseline) for each of the three list types used in Experiment 3: semantic, phonological, and hybrid. There are three important points to note in Table 4. First, recognition for semantic associates and the corresponding critical items were both characterized by a high proportion of "remember" relative to "know" responses. Specifically, collapsing across recall and math test types, compared to their respective baselines (≅.03), both veridical (.43) and false recognition (.32) had a relatively large contribution from "remember" responses [veridical, $F(1,31) = 87.04$, $MSe = .028$, $p < .001$; false, $F(1,31) = 64.45$, $MSe = .021$, $p < .001$]; whereas compared to their respective baselines (≅.13), both veridical (.23) and false recognition (.21) had a relatively small contribution from "know" responses [veridical, $F(1,31) = 13.34$, $MSe = .017$, $p < .001$; false, $F(1,31) = 2.51$, $MSe = .032$, $p = .12$]. Second, in contrast to the findings with lists of semantic associates, recognition for phonological associates and the corresponding critical items were both characterized by relatively more "know" than "remember" responses. That is, collapsing across recall and math test types, compared to their respective baselines (≅.03), both veridical (.16) and false recognition (.11) had a small contribution from "remember" responses [veridical, $F(1,31) = 26.23$, $MSe = .011$, $p < .001$; false, $F(1,31) = 19.24$, $MSe = .006$, $p < .001$]; whereas compared to their respective baselines (≅.13), both veridical (.31) and false recognition (.23) had a slightly larger contribution from "know" responses [veridical, $F(1,31) = 26.33$, $MSe = .022$, $p < .001$; false, $F(1,31) = 5.55$, $MSe = .021$, $p < .03$]. Thus, in general, semantic lists were relatively more likely to influence "remember" responses, whereas phonological lists were relatively more likely to influence "know" responses. Third, for hybrid lists performance appeared to depend on the presence of a preceding recall or math test. Specifically, when preceded by study and recall, recognition of the corresponding critical item was characterized by a high proportion of "remember" relative to "know" responses (.36 to .24, respectively), whereas for hybrid lists preceded by study and math, recognition of the corresponding critical item demonstrated the opposite pattern and was characterized by a high proportion of "know" relative to "remember" responses (.34 to .26, respectively).

A series of ANOVAs were conducted to further explore the influence of list type on "remember–know" judgments. Consistent with the above observations, when hybrid lists were preceded by study and recall, recognition of the non-presented critical item varied as a function of list type for "remember" responses, $F(2,62) = 6.43$, $MSe = .065$, $p < .005$, but not for "know" responses, $F(2,62) = 1.53$, $MSe = .055$, $p = .23$. Moreover, as in the recognition data reported earlier, when preceded by study and recall, false recognition in "remember" responses for hybrid lists (.36) was over-additive with respect to the predicted probability for "remember" responses (.27), although this difference did not reach significance, $F(1,31) = 2.74$, $MSe = .058$, $p = .11$. When hybrid lists were preceded by study and math, recognition of the non-presented critical item varied as a function of list type for both "remember" responses, $F(2,62) = 15.99$, $MSe = .029$, $p < .001$, and "know" responses, $F(2,62) = 4.01$, $MSe = .045$, $p < .03$. Furthermore, when preceded by study and math, false recognition for hybrid lists was over-additive with respect to the predicted average in both "remember" responses (.26 vs .16), $F(1,31) = 6.29$, $MSe = .025$, $p < .02$, and "know" responses (.34 vs .23), $F(1,31) = 7.63$, $MSe = .027$, $p < .01$.

Because there were different levels of recognition performance as a function of list type, test type, and item type, the inferences drawn from the above "remember–know" analyses may be somewhat compromised. To address this possibility, Table 5 presents the proportion of "remember" and "know" responses in each cell as a function of the overall recognition performance in that condition (excluding baseline performance, which was

Table 5
Proportion of remember and know responses conditional on overall recognition performance in each cell

	List Type					
	Semantic		*Phonological*		*Hybrid*	
	Recall	*Math*	*Recall*	*Math*	*Recall*	*Math*
Remember						
Veridical	.68	.62	.33	.37	.65	.49
False	.69	.52	.39	.21	.60	.44
Know						
Veridical	.32	.37	.67	.63	.37	.49
False	.31	.50	.61	.79	.40	.58

Note: Rounding error may prevent the probability of "remember" and "know" responses from summing to equal 1.

characterized primarily by "know" responses for both veridical and false recognition). However, because there were missing cells for some participants, we primarily report these "remember" and "know" proportions at a descriptive level. As shown in Table 5, the general pattern described above held for semantic and phonological lists. Specifically, semantic lists were relatively more likely to influence "remember" responses, whereas phonological lists were relatively more likely to influence "know" responses. As noted earlier, this cross-over pattern occurred for both veridical and false recognition. Turning to the hybrid lists, these lists produced over-additive false recognition in "remember" judgments that were roughly equivalent ($\cong$.07) following earlier recall (predicted vs observed equals .54 vs .60, respectively) and earlier math (predicted vs observed equals .37 vs .44, respectively).

Discussion

The results of Experiment 3 indicated that false recall with hybrid lists was over-additive with respect to the average probability of false recall observed with pure semantic and pure phonological lists. In contrast to Experiment 2, this over-additive effect must have been produced by the mixture of semantic and phonological associates rather than the presence of additional associates, because both pure and hybrid lists contained 16 total associates. A similar over-additive false memory effect was also obtained with hybrid lists in recognition performance. Moreover, after correcting for baseline levels of false recognition, consistent with prior work by Schacter et al. (1997), recognition for semantic associates and false recognition for the corresponding critical items were both characterized by a high proportion of "remember" relative to "know" responses; however, recognition for phonological associates and false recognition for the corresponding critical items were both characterized by a roughly equivalent proportion of "remember" and 'know" responses (see Table 4). Finally, over-additive false recognition with hybrid lists was reflected in "remember " judgments, regardless of whether recognition was preceded by a recall task or a math filler task.

General discussion

The present study yielded converging evidence that the combined presence of semantic and phonological associates produces over-additive influences on false memory performance. This pattern was replicated across three different experimental approaches to testing over-additivity, in both recall and recognition performance, and for "remember–know" judgments. We now turn to a discussion of the implications of these results for understanding the mechanisms that may underlie false memories in the hybrid list technique, including spreading activation and a breakdown in source monitoring.

Creating false memories with hybrid lists of associates: over-additive false memories from converging associative networks

The present results suggest that some additional assumptions must be made with respect to a simple spreading activation account of false memories (see Roediger, Balota, & Watson, 2001). Specifically, this study allows an interesting test of simple additivity of semantic and phonological influences on false memory. For example, the results from Experiment 1 clearly indicated that one finds a much larger effect by adding three phonological associates (.25) to a list of semantic associates than by adding three more semantic associates (.02). Experiment 2 produced a remarkably high level of false recall with hybrid lists (.65) which was over-additive with respect to the predicted sum of false recall observed with pure semantic and pure phonological lists (.48). Experiment 3 extended this pattern to a situation where the total number of associates was equated across lists and also extended the over-additive pattern to recognition performance. In Experiment 3, inconsistent with a simple activation account, false recognition and false recall were over-additive by as much as .20 to .25, respectively. Hence, the results from these three experiments are not consistent with the notion that activation linearly adds across different types of codes to produce false memories, as is predicted by simple associative theories (see Balota & Paul, 1996).

It is interesting to note here that there have been other instances of over-additive effects of semantic and phonological information in memory and lexical processing. For example, using a word generation task, Rubin and Wallace (1989) found that participants who were given meaningful cues such as "a mythical being" did not produce *ghost*; similarly, those who were given the cue "rhymes with ost" failed to produce *ghost*. However, if given the combined cue "a mythical being that rhymes with ost," the participants produced *ghost* every time. Hence, in Rubin and Wallace's study, the independent cues were ineffective in eliciting the targeted response, but the compound cue always did. Consistent with this notion, Balota, Pollatsek, and Rayner (1986) and Becker and Killion (1977), among others, have reported over-additive interactions in eye fixation durations during reading and primed lexical decisions, respectively, when there are multiple constraints from semantic context and phonological information. As discussed below, over-additive effects also occur in speech production (e.g., substituting the word *start* for *stop* based on converging semantic and phonological features). Thus, one of the most the intriguing aspects of the present study is that these over-additive effects were extended to false memory paradigms.

It should also be noted here that the over-additive false memory effects observed with the hybrid lists were not obvious from the start of this research program. Specifically, in a hybrid list, the non-presented critical item is the only item that meets the constraints of the converging semantic and phonological associates (e.g., *dog* is the only word that is semantically related to *cat* and *bark* and phonologically related to *log* and *jog*). This convergence could make a critical item's absence from a study list conspicuous to participants and reduce the incidence of false memories (especially when combined with a warning against guessing in the instructions, as in Experiment 3). Clearly, the results of the present study suggest that this pattern did not occur. In fact, of the 154 participants tested in Experiments 1–3, only 15 participants had to be removed from the analyses because they were *suspected* to have some knowledge of the false memory manipulation, as reflected by a post-experiment question-and-answer session and questionnaire. Thus, although there may have been some transparency to false memory manipulations with the hybrid list technique, the majority of the participants in this study did not detect the manipulation and demonstrated robust false memories with hybrid lists of associates.

In order to better understand the potential influence of noticing the absence of the critical item, we conducted separate analyses on the excluded participants. First, consider the recall performance for the eight excluded participants in Experiment 1 who received hybrid lists of associates. False recall increased from a base rate of .05 in *S* lists, to a rate of .08 in *SR1* lists, to a rate of .12 in *SR2* lists, to a rate of .20 in *SR3* lists for these participants. Although the absolute level of false recall for the excluded participants was reduced compared to the remaining participants, false recall as a function of list type was qualitatively similar to the pattern depicted in Fig. 1, where false recall increased with increasing numbers of rhyming associates. Now consider the recall performance for the four excluded participants in Experiment 2. Specifically, false recall was .13 for pure semantic and pure phonological lists, but increased to .25 for hybrid lists. Unlike the remaining participants in Experiment 2, false recall following hybrid lists of associates appeared to be additive for the excluded participants. However, as in Experiment 1, the excluded participants in Experiment 2 did not have a marked reduction in false recall with hybrid lists of associates relative to pure lists, as one would expect if the hybrid list technique for eliciting false memories was transparent and the omission of critical items was conspicuous. Indeed, for both sets of excluded participants, the absolute level of false recall was reduced, but not eliminated, with some knowledge of false memory manipulations. This last observation is consistent with the finding that individuals can reduce, but not eliminate, false recognition to critical items when given detailed, explicit warnings to avoid false memories at encoding prior to the presentation of pure lists of semantic associates (see Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001; McDermott & Roediger, 1998). Although the data from the excluded participants in the present study certainly are intriguing, some caution is warranted in interpreting these results. Clearly, additional research is necessary to determine (1) if there is an influence of a warning manipulation on false memories elicited by the hybrid list technique, and if a warning does reduce false memories, (2) whether its influence is greater on false memories observed with hybrid lists of associates or pure lists of associates (e.g., Are false memories still over-additive for hybrid lists when a warning is given at encoding?).

Potential mechanisms for over-additive false memories

There are a number of distinct frameworks to accommodate the over-additive influences of semantic and phonological information observed in the present study. First, it is possible that one may be able to modify a simple spreading activation account by assuming that over-additive false memories may be a reflection of distinct lexical representations that are at different points in sensitivity in activation functions. For example, if there are distinct semantic and phonological lexical representations that accumulate activation, it is possible that a mixture of these codes yields the most effective activation pattern (e.g., see Besner, 1990). That is, the activation function in a particular associative network (e.g., semantic, phonological) may be likely to reach a peak sensitivity range with increasing numbers of associates in that network. Hence, there is a diminishing return from additional words within the same domain, because activation has already approached asymptotic levels for that associative representation. Indeed, there is some empirical support for this notion of diminishing returns from additional words within a single domain. Specifically, Robinson and Roediger (1997) found that increasing the number of semantic associates increased false recall of critical items, but not in a strictly linear fashion (although, this pattern may be due, in part, to the low BAS of the last few items in a 15-item list of semantic associates). Furthermore, the results of Experiment 1 replicated the results of Robinson and Roediger because false recall of critical items increased from a baseline probability of about .21 with 10 semantic associates to a probability of about .23 with 13 total semantic associates (where mean BAS for the 10 semantic associates and the three additional semantic associates were roughly equivalent for each 15-item list). Thus, pure lists do not produce a maximal increase in activation, whereas hybrid lists may be in the maximal sensitivity range across independent semantic and phonological associative networks, thereby producing an over-additive influence on memory performance.

However, the results of the present study may place some constraints on an optimal activation account of over-additive false memories. For example, in Experiment 1, false memories were assessed following pure lists of semantic associates as well as hybrid lists of associates that contained the same associates as the semantic lists, plus varying numbers of phonological associates (e.g., 1, 2, or 3 rhymes). As shown in Fig. 1, false recall increased dramatically from .24 with pure semantic lists to .49 with hybrid lists that consisted of the same 10 semantic associates from the pure lists plus three additional phonological associates. Although Experiment 1 was not designed to directly test the additivity of false memories in the hybrid list technique, from an optimal activation perspective, it is unlikely that three phonological associates alone (selected randomly from a pool of potential rhyming associates) could produce such a dramatic change in the functional activation that would double the observed false recall rate. In this light, the results of Experiment 1 suggest that some additional factor may be contributing to the over-additive false memories elicited by hybrid lists of associates relative to pure lists of associates. Furthermore, the results of both Experiment 2 and Experiment 3 produced evidence of over-additive influences on false memories with hybrid lists (with different lists lengths across the two experiments). Hence, although different points within the activation function for semantic and phonological networks may be able to accommodate some aspects of the present results, we believe that such a framework will have difficulty accommodating the entire range of manipulations and patterns of over-additivity observed in the present study.

Although simple additive activation models do not appear to accommodate all of the present results, there is an interactive activation model of speech production that predicts that semantic and phonological information can combine to produce over-additive influences in speech errors, (e.g., substituting the word *start* for *stop* based on converging semantic and phonological features, Dell & O'Seaghdha, 1992). Specifically, a recent model of speech production developed by Dell and colleagues (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997) postulates two stages in lexical access. In the first stage, lemma access, a word is mapped onto its conceptual representation in a semantic network. In the second stage, the lemma is mapped onto the phonological form of the word in a phoneme network. By hypothesizing two stages to lexical access, this model is consistent with the notion that semantic and phonological information are uncorrelated dimensions of language. More importantly, the connections in the model are both top-down and bottom-up, allowing activation to spread from semantic units to word units to phoneme units, and vice-versa. In this way, the combination of top-down and bottom-up activation in the model gives a convergent item (e.g., *dog*) a much better chance of occurring as an error in the presence of semantic (e.g., *cat*) and phonological neighbors (e.g., *log*), relative to a purely semantic (e.g., *bark*) or purely phonological error (e.g., *jog*). It is possible that both bottom-up and top-down activation processes combined during list study or during retrieval to surpass the threshold for output. In this way, the present results are consistent with the interactive activation model of speech production developed by Dell and his colleagues. Hence, it may be that some false memories in the DRM paradigm and in the hybrid list technique are simply memory production errors, obeying the same laws as speech production errors.

A third approach to accommodate the present results would be to emphasize the breakdowns in source monitoring that may occur in hybrid lists (see Roediger, Watson, McDermott, & Gallo, 2001, for an activation-monitoring framework of false memories in the DRM paradigm). As noted in the Introduction, the bulk of the DRM literature has shown that individuals falsely remember critical items following lists of lexical/semantic associates, even though there is relatively little perceptual support for critical items in semantic lists (see McDermott, 1997). Pure lists of semantic associates may create a situation where studied items and critical items are highly confusable in participants' internal lexical/semantic networks; however, studied items and critical items may be easier to discriminate on the basis of orthographic/phonological information in the external study environment (see Norman & Schacter, 1997). In the present study, as additional phonological associates are presented in the context of a hybrid list, it may be increasingly difficult for participants to determine whether the critical item was presented or whether items that looked, sounded, or meant something similar were presented. For example, participants may not be able to remember if *dog* was presented at the beginning of a list, or whether it was really *cat* and *log*, because the presentation of *log* provides perceptual support for the conceptually activated stimulus *dog*. In this way, hybrid lists provide perceptual support for critical items in addition to conceptual support, which may make studied items and critical items especially difficult to discriminate on semantic, orthographic, and phonological dimensions of language.

A recent study by Watson et al. (2001) may provide some leverage in testing predictions from an activation-monitoring account of false memories. Specifically, using 12-item semantic, phonological, and hybrid lists of associates, Watson et al. (2001) found that young adults, older adults, and Alzheimer patients produced over-additive false recall of critical items with hybrid lists. Given that spreading activation in underlying associative networks is thought to be relatively intact with age (Balota, Watson, Duchek, & Ferraro, 1999b), whereas attentional control/source monitoring is thought to be relatively impaired (Balota et al., 1999a; Spieler, Balota, & Faust, 1996), one might expect a greater over-additive effect with hybrid lists in older adults compared to young adults. Consistent with this prediction, when false recall was considered as a proportion of veridical recall, older adults did have increased over-additive false recall with hybrid lists relative to young adults (see Fig. 3 in Watson et al., 2001). The results of the present study complemented and extended the Watson et al. (2001) findings (1) by replicating their over-additive pattern in false memories across three different experimental approaches to testing over-additivity (in both recall and recognition performance and for "remember–know" metamemory judgments), and (2) by developing a larger set of materials (see the appendix) for researchers to use when investigating potential mechanisms underlying false memories elicited by semantic, phonological, and hybrid lists of associates (e.g., see Chan et al., 2003, for evidence that pure semantic and pure phonological lists can be used to dissociate levels-of-processing and item-specific/relational processing accounts of false memories in the DRM paradigm). Ultimately, additional studies that directly compare associative priming effects for critical items in word recognition tasks with false memories for the same critical items in episodic memory tasks may be useful in determining (1) the underlying activation functions for pure and hybrid lists of associates, and (2) what role strategic processes such as source monitoring may play in modulating these activation functions during encoding and retrieval.[7]

[7] Although we have only discussed three possible mechanisms underlying over-additive false memories with hybrid lists of associates, there are other models which may help explain these findings. For example, as noted by an anonymous reviewer, Murnane's ICE (Item–Context–Ensemble) theory may be a useful framework for explaining the robust false memories elicited by hybrid lists of associates (e.g., see Murnane, Phelps, & Malmberg, 1999). Specifically, according to this theory, item-specific information and associative, contextual information combine to form an ensemble that influences episodic memory performance. Applying this framework to the present results, pure semantic lists provide contextual support for conceptually driven false memories, whereas orthographic and phonological information may be used to discriminate studied items from non-presented, semantically related critical items. Furthermore, pure phonological lists provide contextual support for perceptually driven false memories, whereas lexical/semantic information may be used to discriminate studied items from non-presented, phonologically related critical items. By combining lexical/semantic and orthographic/phonological associates in a hybrid list, item-specific information and associative, contextual information are fully integrated with respect to the critical item. Hence, in hybrid lists of associates, item-specific information cannot be used as effectively as in pure lists to reduce false memories, which may contribute to the over-additive false memories observed in hybrid lists of associates.

Using the hybrid list technique to dissociate semantic and phonological networks

One of the fundamental assumptions of both an activation-monitoring account and an interactive activation account of the over-additive false memories elicited by the hybrid list technique is the relative independence of semantic and phonological associative networks. That is, meaning and pronunciation are uncorrelated dimensions of language (i.e., the linguistic universal of arbitrariness). If this is the case, in addition to producing over-additive false memories, the hybrid list technique may provide evidence that dissociates the influence of these associative networks. For example, consider the metamemory judgment results for the semantic and phonological associates and the corresponding critical items. Specifically, consistent with Roediger and McDermott (1995), semantic associates and the corresponding critical items were characterized by a high proportion of "remember" relative to "know" responses, suggesting that participants remembered something specific about the context of the majority of these items, even if, as in the case of critical items, they were not presented. In contrast, phonological associates and the corresponding critical items were characterized by a roughly equivalent proportion of "remember" and "know" responses, suggesting that participants were less likely to remember something specific about the context of the items (also see Schacter et al., 1997). Thus, semantic lists were relatively more likely to influence "remember" judgments and phonological lists were relatively more likely to influence "know" judgments (also see Rajaram, 1993, for evidence that conceptual variables influence "remember" responses but not "know" responses in a levels-of-processing framework). Consistent with these findings, recent fMRI research on conceptual and perceptual processing by McDermott, Petersen, Watson, and Ojemann (2003) suggests that attending to meaningful relations among lists of lexical/semantic associates preferentially activates distinct regions of the brain compared to attending to sound relations among lists of orthographic/phonological associates (and vice-versa). Finally, consider false recognition with hybrid lists of associates. These lists produced an over-additive pattern in "remember" judgments following earlier recall and earlier math performance. This pattern is consistent with the notion that hybrid lists provide robust conceptual and perceptual support for critical items relative to pure lists of semantic and phonological associates. Moreover, the over-additive influence on "remember" judgments also suggests that the critical item may have been consciously accessed during encoding. Further support for this idea comes from false recognition data by Gallo and Roediger (2002) who found that semantic lists high in backward associate strength (to the critical item) increased "remember" judgments relative to lists low in backward associative strength.

Conclusions

The present study combined semantic and phonological codes to investigate how multiple sources of associative information influence false memories. Consistent with Watson et al. (2001), the results of three experiments clearly indicated that there are exaggerated effects when there is convergence from both conceptual and perceptual processing domains. In addition, the influence of semantic and phonological processing networks was dissociated in episodic memory performance, including veridical and false recall, veridical and false recognition, and "remember–know" metamemory judgments. Taken together, these findings suggest that individuals are particularly susceptible to false memories when there is a convergence of relatively independent dimensions of stimuli (meaning and pronunciation) on non-presented events.

Acknowledgments

This research was supported by National Institute on Aging Grant P01 AGO3991. We thank Kefong Lee and Rachel Denison, who provided help with various aspects of this project. We thank Keith Hutchison, Jes Logan, Kathleen McDermott, and Jason Chan for providing helpful discussion at different points in the project. We also thank Dawn McBride and three anonymous reviewers for helpful comments on earlier versions of the manuscript.

Appendix: Lists of studied items for each critical item used in Experiment 3

BAD		BALL		BEER		BLACK	
good	had	bounce	doll	drunk	leer	white	mack
rotten	lad	throw	bile	keg	peer	gray	block
harmful	bat	basket	bail	pub	tear	tar	blank
worse	bag	bowling	balk	suds	rear	bruise	lack
villain	bud	golf	wall	liquor	seer	brown	sack
severe	band	play	fall	booze	gear	oil	smack
trouble	dad	tennis	bald	alcohol	bill	tuxedo	track
awful	bide	soccer	pall	Bud	deer	dark	pack
terrible	bid	round	tall	bar	boar	prejudice	snack
evil	pad	catch	bill	bottle	beard	minority	rack
corrupt	ad	pitch	bell	wine	hear	coffee	flack
horrible	bed	moth	all	mug	fear	color	slack
nasty	ban	bat	boil	barrel	year	Africa	bleak
attitude	tad	kick	bull	drink	bear	coal	back
mood	sad	racket	gall	can	veer	soul	hack
punish	fad	hit	hall	cooler	ear	race	plaque

BREAD		CAR		CHAIR		COLD	
rye	bled	auto	char	sit	pair	chill	code
loaf	bride	drive	call	couch	share	hot	called
crust	braid	engine	care	rocking	char	warm	fold
wheat	read	wreck	are	swivel	air	sneeze	sold
butter	broad	garage	card	cushion	scare	shiver	culled
crumb	bed	motor	carp	seat	check	Arctic	chord
garlic	thread	van	cot	recliner	lair	ice cream	scold
muffin	tread	truck	core	wicker	hair	chilly	bold
dough	brad	crash	par	Lazyboy	their	freezer	hold
toast	pled	accident	scar	table	tear	frigid	coiled
flour	wed	trunk	cart	stool	cherry	heat	colt
Wonder	breed	tire	far	furniture	cheer	ice	old
bun	breadth	mechanic	bar	sofa	stair	frost	polled
baked	fled	vehicle	carve	rocker	fair	freeze	gold
biscuit	head	tow	cough	desk	care	winter	told
roll	dread	gas	tar	bench	chore	snow	coal

References

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., & Yerys, B. E. (1999a). Veridical and false memories in healthy older adults and in Dementia of the Alzheimer's type. *Cognitive Neuropsychology, 16*, 361–384.

Appendix (*continued*)

DOG		FACE		FAT		FLAG	
hound	log	mouth	fake	thin	fate	American	slag
puppy	dodge	expression	vase	obese	that	banner	flab
bite	dug	nose	fuss	large	sat	pledge	brag
mutt	hog	eyes	faith	weight	foot	wave	wag
pet	bog	frown	lace	calorie	fact	allegiance	flak
beware	doff	wrinkle	fail	slim	cat	country	sag
bone	daub	makeup	fain	pudgy	feat	stars	nag
tail	cog	cheek	ace	diet	fit	USA	snag
cat	dock	head	case	slender	bat	pole	bag
animal	dawn	mask	fate	wide	pat	stripes	crag
paw	fog	moustache	fame	cheek	fan	freedom	flat
poodle	dig	beard	race	skinny	fast	nation	lag
flea	doll	chin	base	lean	hat	pennant	gag
bark	frog	lips	faze	plump	fought	salute	flog
Lassie	jog	shave	fade	chubby	flat	symbol	drag
vet	dot	smile	pace	huge	at	checkered	rag

GOD		GLASS		GUN		HAND	
lord	pod	bottle	class	pistol	gown	glove	land
holy	gone	lens	grass	shot	bun	finger	sand
heaven	goad	shatter	blass	holster	nun	shake	hound
bible	odd	prism	lass	rifle	gush	palm	panned
bless	tod	mirror	glaze	bullet	one	thumb	stand
angel	good	hour	sass	hunt	ton	wave	hanged
sin	sod	crystal	bass	military	gut	grip	fanned
faith	wad	jar	glance	powder	sun	foot	canned
church	guide	pane	mass	shoot	run	fist	band
Jesus	nod	fragile	brass	trigger	goon	mitten	grand
religion	gob	mug	crass	murder	gain	wash	honed
pray	gad	looking	gloss	aim	gum	hold	hind
devil	gall	shard	glad	bang	pun	knuckle	tanned
deity	rod	cup	plass	cannon	fun	wrist	and
Christ	cod	break	pass	revolver	done	arm	had
worship	got	window	gas	weapon	gone	clap	brand

HARD		HATE		KILL		LAW	
rigid	bard	dislike	rate	slay	skill	rights	raw
difficult	hark	love	wait	suicide	cull	attorney	paw
easy	harm	hostility	hail	violence	kid	enforce	chaw
work	lard	anger	hot	hunt	hill	criminal	lawn
cement	charred	detest	fate	shoot	fill	lawyer	lock
concrete	scarred	resent	haste	stab	chill	court	claw
stiff	hoard	fear	height	attack	sill	government	flaw
tough	hired	jealousy	date	homicide	kilt	regulation	log
rock	sparred	envy	gate	destroy	call	legal	lay
simple	heart	despise	hay	shot	coil	officer	saw
complex	harp	abhor	bait	smother	till	rules	gnaw
firm	starred	war	late	poison	pill	justice	low
solid	tarred	enemy	hat	assassin	kick	legislation	lot
soft	yard	loathe	hit	murder	kit	amendment	awe
rough	card	disgust	heat	deadly	keel	police	slaw
coarse	herd	like	ate	choke	ill	order	loss

MAIL		MAN		PEN		RAIN	
stamp	meal	woman	can	ink	pan	umbrella	train
deliver	nail	guy	moon	paper	then	drench	main
receive	mate	sir	main	marker	hen	weather	ran
bills	mile	boss	fan	eraser	ken	hail	wren

Appendix (*continued*)

letters	hail	super	tan	pencil	pawn	cloud	pain
send	make	lady	pan	writing	pain	dew	rave
fax	mall	person	mean	notebook	fen	pour	raise
express	sail	fellow	map	Bic	peg	storm	brain
post	veil	mister	van	point	when	thunder	bane
zip	mill	bachelor	ran	mark	ben	wind	raid
address	mole	uncle	mat	write	pine	puddle	rate
envelope	maid	con	mad	scribble	pun	acid	range
package	may	macho	ban	pal	yen	mist	wane
UPS	ail	handsome	mine	quill	ten	lightning	lane
telegram	gale	gentleman	moan	fountain	pet	sunshine	vain
junk	mull	male	an	pad	pent	flood	gain
RIGHT		SICK		SLEEP		SLOW	
correct	tight	healthy	sock	bed	sweep	quick	mow
perfect	rye	ill	sink	rest	steep	fast	crow
equal	rife	flu	lick	yawn	sleet	snail	slope
accurate	night	nausea	sake	pillow	slop	hesitant	slaw
fair	bright	cancer	soak	snooze	heap	brisk	owe
justify	rile	cough	kick	awake	weep	swift	snow
left	ripe	virus	six	nap	seep	molasses	blow
turn	bite	disease	suck	dream	sleek	lazy	throw
angle	rat	medicine	silk	tired	slope	cautious	row
answer	rot	doctor	sack	pajamas	bleep	lethargic	flow
mistake	white	fever	stick	snore	slip	speed	slew
wrong	rice	hospital	thick	doze	slap	hurry	hoe
truth	ride	germ	seek	drowsy	leap	sluggish	show
ethics	light	clinic	slick	coma	cheap	turtle	sew
direction	writhe	vomit	tick	wake	sleeve	rapid	glow
proper	rate	well	sip	slumber	sloop	delay	low
SMELL		SMOKE		SNAKE		SWEET	
odor	bell	fire	poke	viper	brake	honey	beat
cologne	swell	nicotine	joke	lizard	quake	bitter	heat
sniff	spell	cigar	cloak	slither	snack	nice	skeet
stench	tell	pot	smirk	serpent	sake	ice cream	street
scent	hell	pipe	stroke	deadly	ache	sugar	swat
nose	smile	chimney	oak	hiss	snuck	tart	wheat
deodorant	yell	fumes	smote	reptile	shake	taste	feet
aroma	jell	cigarette	spoke	cobra	flake	fudge	meet
skunk	small	ashtray	choke	fangs	lake	candy	sleet
fragrance	fell	Marlboro	bloke	poison	make	syrup	seat
dirty	knell	marijuana	woke	venom	sneak	kind	sweep
sense	sell	smog	smack	slimy	stake	chocolate	fleet
perfume	shell	habit	stoke	bite	rake	dessert	sheet
stink	well	tobacco	yoke	python	snail	sour	sweat
foul	dell	puff	smock	worm	take	frosting	treat
whiff	smelt	inhale	soak	rattle	wake	salty	tweet
TEST		TOP		TRASH		WET	
quiz	zest	bottom	mop	garbage	gash	slippery	vet
final	pest	peak	stop	waste	slash	damp	watt
study	tossed	hill	tap	dumpster	track	paint	wheat
evaluate	west	over	tup	junk	brash	splash	pet
experiment	chest	roof	chop	refuse	flash	dry	west
essay	tent	summit	bop	Hefty	ash	humid	bet
stress	toast	pinnacle	tock	litter	stash	water	wed
screen	crest	zenith	cop	sewage	trap	dripping	well
score	fest	apex	hop	scraps	lash	soak	net

Appendix (*continued*)

exam	best	spin	tape	dump	rash	moist	let
fail	text	above	taupe	rubbish	mash	saturate	welt
lab	taste	ceiling	pop	landfill	thrash	sponge	wit
tube	vest	tip	type	can	bash	towel	wait
pass	hest	lid	tot	pile	dash	slick	get
grade	rest	mountain	sop	bag	crash	soggy	yet
analysis	guest	best	whop	recycle	clash	douse	wear

Note: For each critical item, semantic associates are presented in the left column and phonological associates are presented in the right column. Hybrid lists are created by interleaving semantic associates from odd-numbered serial positions with phonological associates from even-numbered serial positions, or vice-versa, for each critical item, as shown in Table 3.

Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 827–845.

Balota, D. A., Pollatsek, A., & Rayner, K. (1986). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*, 364–390.

Balota, D. A., Watson, J. M., Duchek, J. M., & Ferraro, F. R. (1999b). Cross-modal semantic and homograph priming in healthy young, healthy old, and in Alzheimer's disease individuals. *Journal of the International Neuropsychological Society, 5*, 626–640.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 389–401.

Besner, D. (1990). Does the reading system need a lexicon. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 73–96). Hillsdale, NJ: Erlbaum.

Carnegie-Mellon Semantic Rhyming Dictionary [On-line]. Available: www.rhymezone.com.

Chan, J. C. K., McDermott, K. B., Watson, J. M., & Gallo, D. A. (2003). The importance of relational processing in producing false memories (submitted).

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287–314.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801–838.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Memory & Cognition, 4*, 271–276.

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*, 469–497.

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review, 8*, 579–586.

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

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1–36.

McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through the presentation of associates. *Psychonomic Bulletin & Review, 4*, 582–586.

McDermott, K. B., Petersen, S., Watson, J. M., & Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia, 41*, 293–303.

McDermott, K. B., & Roediger, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39*, 508–520.

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language, 45*, 160–176.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General, 128*, 403–415.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). The University of South Florida word association, rhyme, and word fragment norms. Unpublished manuscript, University of South Florida, Tampa.

Norman, K., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838–848.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89–102.

Redfield, B. (1986). *Capricorn rhyming dictionary*. New York, NY: Putnam Publishing Group.

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*, 231–237.

Roediger, H. L., Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant

(Eds.), *The nature of remembering: essays in honor of Robert G Crowder* (pp. 95–115). Washington, DC: American Psychological Association Press.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Roediger, H. L., McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187–246). Hove, Sussex: Psychological Press.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385–407.

Rubin, D. C., & Wallace, W. T. (1989). Rhyme and reason: Analyses of dual retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 698–709.

Russell, W. A., & Jenkins, J. J. (1954). The complete Minnesota norms for responses to 100 words from the Kent–Rosanoff Word Association Test. Technical Report No. 11, Contract N8 ONR 66216, Office of Naval Research, University of Minnesota, Minneapolis.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1–24.

Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: Conceptual and perceptual false recognition. *Neuropsychology, 11*, 331–342.

Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language, 40*, 83–108.

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in younger adults, healthy older adults and individuals with senile Dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 461–479.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*, 494–500.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26*, 1–12.

Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with Dementia of the Alzheimer type. *Neuropsychology, 2*, 254–267.