Original Article

QJEP

Short-term false memories vary as a function of list type

Quarterly Journal of Experimental Psychology
2019, Vol. 72(12) 2726–2741

DOI: 10.1177/1747021819859880
qjep.sagepub.com
SAGE

Dawn M McBride[1], Jennifer H Coane[2], Shuofeng Xu[2], Yi Feng[2] and Zhichun Yu[2]

Abstract
False memories have primarily been investigated at long-term delays in the Deese–Roediger–McDermott (DRM) procedure, but a few studies have reported meaning-based false memories at delays as short as 1–4 s. The current study further investigated the processes that contribute to short-term false memories with semantic and phonological lists (Experiment 1) and hybrid lists containing items of each type (Experiment 2). In Experiment 1, more false memories were found for phonological than for semantic lists. In Experiment 2, an asymmetrical hyper-additive effect was found such that including one or two phonological associates in pure semantic lists yielded a robust increase in false alarms, whereas including semantic associates in pure phonological lists did not affect false alarms. These results are more consistent with the activation–monitoring account of false memory creation than with fuzzy trace theory that has not typically been referenced when describing phonological false memories.

Keywords
False memories; activation; source monitoring; short-term memory; phonological and semantic errors

Received: 19 March 2019; revised: 28 May 2019; accepted: 29 May 2019

False memories are of interest to memory researchers because they can occur automatically and often without awareness, highlighting normal processes in human memory. In the past few decades, this area of research has benefitted from use of the Deese–Roediger–McDermott (DRM) (Deese, 1959; Roediger & McDermott, 1995) paradigm, where lists of items that are semantically and/or associatively related to a non-presented lure item are studied by participants.[1] For example, participants study a list containing the words *awake, doze, dream, pillow, snore*, and so on, all related to the theme *sleep*. On a subsequent recall or recognition test, *sleep* is often freely recalled with other studied items from this list or endorsed as a studied item in a recognition test. This procedure has been used countless times to consistently produce false memories for the theme critical items (CIs), allowing researchers to test ideas about how false memories are created through normal memory processes (e.g., Gallo, 2010).

One common explanation for the creation of false memories for CIs in the DRM procedure relies on a combination of activation and monitoring processes. A number of researchers (e.g., Roediger, Balota, & Watson, 2001) have suggested that when the associated list items are studied, the CI is activated in semantic memory. With several list items presented that are all associated with the CI, the activation of the CI item strengthens, thus increasing its familiarity or accessibility. The activation process is fast-acting and does not require intentional encoding or directed attention (Dodd & MacLeod, 2004; Pérez-Mata, Don Read, & Diges, 2002). Later, when participants attempt to retrieve the list items, a source monitoring error occurs (Johnson, Hashtroudi, & Lindsay, 1993), where they mistake this activation for a presentation of the CI in the list. The presentation of list items in the recognition test prior to the CI can also reactivate the CI, increasing the likelihood of a source error (e.g., Coane & McBride, 2006), a finding consistent with the idea that the activation process occurs rapidly and involuntarily.

Additional evidence consistent with the activation–monitoring account of DRM procedure false memories comes from studies manipulating the types of lists presented. For example, lists phonologically and/or

[1]Department of Psychology, Illinois State University, Normal, IL, USA
[2]Colby College, Waterville, ME, USA

Corresponding author:
Dawn M McBride, Department of Psychology, Illinois State University, Campus Box 4620, Normal, IL 61790-4620, USA.
Email: dmcbride@ilstu.edu

orthographically related to the CI (e.g., *sip, gin, sit, been*, and *pin*, related to the CI *sin*) have also shown false memories for the CIs (e.g., Sommers & Lewis, 1999), suggesting that semantic associations are not the only types of connections that can be activated during list presentation. Phonological and orthographic neighbours (i.e., items that differ by one or more phonemes or graphemes from a CI) also activate related items, although the similarity is in terms of shared surface features instead of meaning (e.g., Neighbourhood Activation Model [NAM]; Luce & Pisoni, 1998). For the sake of brevity, we will refer to lists that converge on the CI in terms of shared surface features as "phonologically related" or "phonological lists."

Watson, Balota, and Roediger (2003) further showed that pure semantic and phonological lists produce fewer false memories for the CIs than hybrid lists containing both semantically and phonologically related items to the CIs (e.g., list items of *hound, puppy, log*, and *dot* for the CI *dog*). Their results showed hyper-additivity (i.e., higher false alarm rates to CIs with hybrid lists than with pure lists of one type of the same length) in false memories created by the hybrid lists when compared with pure lists of the same list length. This suggests that convergence of activation from multiple sources increased the accessibility or familiarity of the CI item or disrupted the functioning of monitoring processes by reducing the probability that participants could rely on the absence of diagnostic information (e.g., rejecting the CI because no studied items activated the specific phonological or semantic code).

Finley, Sungkhasettee, Roediger, and Balota (2017) recently replicated Watson et al.'s (2003) findings with all possible list compositions of the hybrid lists to examine the activation function for additional items of each type in the list. Lists of 16 items were presented at study. List composition varied from pure lists (16 semantically related items or 16 phonologically related items) to evenly divided hybrid lists (eight semantic and eight phonological items) with all possible combinations between these points tested (i.e., 15 semantic items and one phonological item, one semantic item and 15 phonological items, etc.). They found hyper-additivity with higher CI recall for hybrid than for pure lists and symmetrical contributions of semantic and phonological items. Interestingly, the hyper-additive effect was most pronounced when one or two items of a different type were added (i.e., the 15-1 and 14-2 conditions), with diminishing returns observed as lists became more balanced. This could be due to the combination of two types of associations in the lists—with two types of information to consider, it is more difficult to reject the CIs using a recollection-rejection strategy.

In addition to results with hybrid lists, Coane, McBride, Termonen, and Cutting (2016) found more false memories for lists with both associative and categorical associations (i.e., *wolf, coyote, hound, fox*, etc.—list items that are both thematically associated with and in the same category as the CI *dog*) than with pure associative lists that contained items that did not share categorical features with the CI items (i.e., list items *pound, bark, beware, collar*, etc. for the CI *dog*). Across two experiments, false recognition of CIs was higher for the associative plus categorical lists than for the pure associative lists, suggesting that shared categorical features contribute additional activation of the list theme beyond activation from the semantic associations. These studies show that different list aspects, specifically, different types of relations between studied items and CIs, can make independent contributions to the false memory effects in the DRM paradigm.

Despite multiple studies showing differential effects of list aspects, it is unclear at what point in the time course of the activation function the contributions from different sources influence the creation of false memories. Although phonologically related lists can create false memories at long-term delayed tests (as noted above), long-term memory is more sensitive to meaning-based influences (e.g., Norris, 2017), whereas short-term memory is more sensitive to phonological/surface similarity (e.g., Nairne & Neath, 2013). Yet, phonologically related lists have not been tested at delays relying on short-term memory processes. Thus, although both semantic and phonological relations clearly contribute to false memories for CIs at the long term, it is unclear what the relative contributions of the different sources of activation are at different retrieval delays. If specific list features (semantic associations, phonological associations) affect false memory creation at short-term delays to different degrees, this will highlight the features that are most highly activated during short-term retrieval and provide further insight into the time course of false memory formation.

Short-term false memories

The majority of studies that have used the DRM procedure to examine false memories have tested memory at delays long enough to be considered long-term memory (i.e., delays from several minutes to hours). However, according to the activation–monitoring account, testing memory after short-term delays, specifically delays of 1 min or less, should reduce or eliminate false memories because veridical recall is negatively correlated with false recall (Roediger, Watson, McDermott, & Gallo, 2001). Furthermore, veridical memory should improve with shorter delays and shorter lists, in part due to improved source monitoring (Frost, Ingraham, & Wilson, 2002; Murdock, 1961; Thapar & McDermott, 2001, but see Sugrue, Strange, & Hayne, 2009). The phonological primacy of short-term/working memory is well established based on findings of phonological similarity effects (e.g., Baddeley, 1964; Conrad, 1963; Wickelgren, 1965) and articulatory suppression effects (e.g., Murray, 1967), suggesting that activation of semantically related lists should not produce robust false memories for CIs in short-term tasks—if false memories occur on short-term tasks, they should be weak. Thus, although activation effects, which occur automatically and rapidly (e.g., Neely, 1977), would increase the accessibility of the CI, the conditions specific to short-term testing should allow participants to reject the CI by relying on controlled monitoring.

However, a number of studies have shown short-term false memories with semantically themed lists. Coane, McBride, Raulerson, and Jordan (2007) found that with lists of five or seven semantically related items, CI probes were falsely recognised with rates at or above 20% with only a 750-ms delay after list presentation. CIs also took longer to be correctly rejected than weakly related unstudied probes. Similar results were reported by Reuter-Lorenz and colleagues (Atkins & Reuter-Lorenz, 2008; Flegal, Atkins, & Reuter-Lorenz, 2010; Flegal & Reuter-Lorenz, 2014). In addition, Olszewska, Reuter-Lorenz, Munier, and Bendler (2015) found short-term false memories for semantic lists when lists were presented in auditory modality, although the effect was weaker than in visual modality. They argued that, due to the primacy of phonological coding in short-term memory (e.g., Nairne, 1990, 2002), the stronger phonological traces of auditorily presented lists made it easier to reject the CIs than for visually presented lists, which did not provide as strong a phonological code.

Abadie and Camos (2018) tested the underlying working memory processes contributing to short-term false memories in four experiments that examined the importance of rehearsal maintenance and attentional refreshing (i.e., an attention-based mechanism that refers to thinking back to recently processed information) on short- and long-term false memories.[2] Their results showed that false memories in semantic lists were evident only when rehearsal maintenance of studied items was blocked with articulatory suppression. When rehearsal processes were not blocked, false memories were significantly reduced in short-term tests. They concluded that working memory mechanisms, specifically refreshing of thematic representations, contribute to semantic long-term false memories, but not to semantic short-term false memories. They suggested that engaging in working memory rehearsal prevented false memories at the short-term through retrieval of physical/surface traces of list items, allowing the rejection of the CIs. However, it is important to note that Coane et al. (2007) did obtain semantic false memories at the short-term even under conditions in which rehearsal was not blocked, although there were procedural differences such as list length, retention interval, presentation rate, and list type that might account for such discrepancies. Thus, there may be conditions under which rehearsal does not prevent semantic false memories at short-term delays.

Abadie and Camos' (2018) conclusions are in line with another account of semantically based false memories, fuzzy trace theory (Brainerd & Reyna, 1998, 2005). Fuzzy trace theory suggests that two types of memory traces can be accessed in retrieval—verbatim (i.e., based on perceptual features of encoded information) and gist (i.e., based on conceptual or semantic features of encoded information) traces. Gist and verbatim traces can be accessed for studied items, and will support correct recognition and recall, but gist is the cause of false alarms to and false recall of CIs. Because gist is similar to that for studied items, retrieval of the gist when evaluating a CI as a studied item causes it to be falsely remembered as a list item. When verbatim memory is high, it can counteract the effects of gist-based processing through a process referred to as recollection-rejection, where participants can use verbatim traces of encoded items to reject similar foils. This is one possible interpretation of Olszewska et al.'s (2015) finding of fewer false alarms to CIs for auditorily presented lists than for visually presented lists in short-term tests, because the auditory presentation allowed for stronger maintenance of verbatim phonological information that discriminated between studied and non-studied items. However, fuzzy trace theory has difficulty explaining phonologically based false memories because gist, which is the main cause of false memories according to the theory, is assumed to be meaning-based. We note that some researchers (e.g., Budson, Sullivan, Daffner, & Schacter, 2003; Holliday & Weekes, 2006) have proposed that phonological false memories could be supported by *phonological gist*, defined as thematic relations between items that are determined by phonological similarity at the sublexical level. However, the original proponents of fuzzy trace theory define gist as "vague, meaning-based memory representations" (Reyna, Corbin, Weldon, & Brainerd, 2018, p. 1). Thus, in its more common usage, gist is assumed to be based on meaning-based relations between list items and CIs.

Furthermore, all published studies using short-term delays to date have only examined semantic false memories—phonological lists have yet to be tested at short-term delays in the DRM procedure. If phonological lists create false memories at the short term, it shows that meaning-based gist is not the primary process driving the effect. In addition, Olszewska et al.'s (2015) conclusion that reliance on stronger phonological codes for auditorily presented lists led to the lower false memory rates suggests that false alarms to CIs for phonological lists should be higher than those for semantic lists. The verbatim similarity of phonological list items to the CI should make recollection-rejection less likely than for semantic lists because the high rate of similarity in terms of phonetic features should result in high levels of activation of the same features in the CI. When using semantic lists, phonological information, or more precisely, the lack of familiarity of specific phonemes, can be used to reject a CI (cf. Watson et al., 2003). Thus, if verbatim information is critical for reducing false memory in the short-term, when such information does not discriminate between studied items and CIs, as in the case of phonological lists, such a reduction should not occur and, in fact, false memories should increase. We tested this prediction in two experiments.

The current study

In the current study, we compared short-term false memories for phonological lists with those for semantic lists in the DRM paradigm. Experiment 1 provided a direct comparison of pure semantic (e.g., *paw, bark, bite, flea*) and phonological (e.g., *bog, dot, dodge, frog*) lists, where lists of each type were used for the same CIs (e.g., *dog*). With phonological lists, we expected more false memories at short-term delays than for semantic lists, because phonological activation should contribute to source errors, where similar phonological codes are more easily confused in immediate tests (Penney, 1975, 1989). This prediction is consistent with Nairne's (1990, 2002) feature model, where similarity in features (particularly phonological features) contributes to retrieval errors, and highlights the unique contribution of phonological information to short-term false memories. It is also consistent with the activation–monitoring account because the activation of similar phonological codes would contribute to source monitoring errors in the immediate test. Furthermore, given we did not prevent rehearsal, this result would impose boundary conditions on Abadie and Camos' (2018) conclusion that rehearsal of verbatim traces over the short term reduces false memories. If we find robust false memories even when participants are allowed to rehearse—albeit briefly, given the very short retention interval—this would suggest that even high levels of verbatim traces (which would be maintained in an un-filled retention interval) may not always effectively counteract the increased familiarity of the CI or gist-based responding due to the phonological similarity of the items.

Experiment 1

Method

Participants. Participants were recruited from Colby College ($N=81$) and Illinois State University ($N=29$), and they participated in exchange for course credit. Data from nine participants were not included in the final analysis either because they were non-native speakers of English or because of experimenter errors. Sample size was determined based on the effect size estimated from the set size five conditions in Coane et al. (2007). An analysis in G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a sample size of 72 was the minimum required to achieve power of at least 80% in the current study. The experimental protocol was approved by the Institutional Review Boards (IRB) at Colby College and at Illinois State University.

Design. The design of Experiment 1 included two factors: list type (semantic vs. phonological) and recognition test probe type (studied list item, CIs, and non-studied items). Both factors were tested within subjects. Figure 1 provides a schematic of the design. As the figure shows, for each list presentation, six items were selected at random from a set of eight items related to the CI (see the Online Supplementary Material for items included in each set). Lists related to each CI were presented once per block across three total blocks. A different probe type was presented in each block. Data analyses were conducted on proportion hits to list item probes and false alarms to CIs and non-studied items. We also examined reaction times (RTs) to correct responses for all probe types and false alarms to CIs and non-studied items.

Materials. The phonological and semantic lists used in the present study were selected from Watson et al. (2003). Two lists were developed from each of the 36 CIs such that each CI had one eight-item phonological associative pool and one eight-item semantic associative pool. The associates were chosen with a goal of matching both associative pools on word length. However, because the CIs were monosyllabic, and many semantic associates were not, we were ultimately unsuccessful in creating length-matched semantic and phonological associative pools, $t(35)=4.26$, standard error (SE)$=.15$, $p<.001$, $d=.70$. It is worth noting that the numerical discrepancy was relatively small, as the semantic associates ($M=4.88$) were on average longer than phonological associates ($M=4.26$) by only .63 letters. Six additional standard DRM lists were used as practice trials.

Procedure. Participants were tested either individually or in small groups of up to three participants at individual work stations. The experiment was administered through E-Prime (Schneider, Eschman, & Zuccolotto, 2012). On each trial, a six-item list was presented immediately followed by a fixation cross, and then a single recognition probe item. Each participant was given six practice trials to familiarise themselves with the experimental paradigm. For the experimental trials, each participant studied 108 six-item lists of words (i.e., memory sets). Half of the sets were phonologically related, consisting of only phonological associates to a CI, and the other half were semantically related, consisting of only semantic associates to a CI. Each word in a set was presented on the screen for 750 ms in white text on a black background followed by an inter-stimulus interval (ISI) of 250 ms. After the sixth item in each set was presented, a fixation cross appeared for 750 ms. After the fixation cross disappeared, a single-item recognition probe was presented, and participants were asked to make their recognition decision. They were instructed to press "L" to indicate an *old* response and "A" to indicate a *new* response. Both speed and accuracy were emphasised. There were three types of probes. The CI

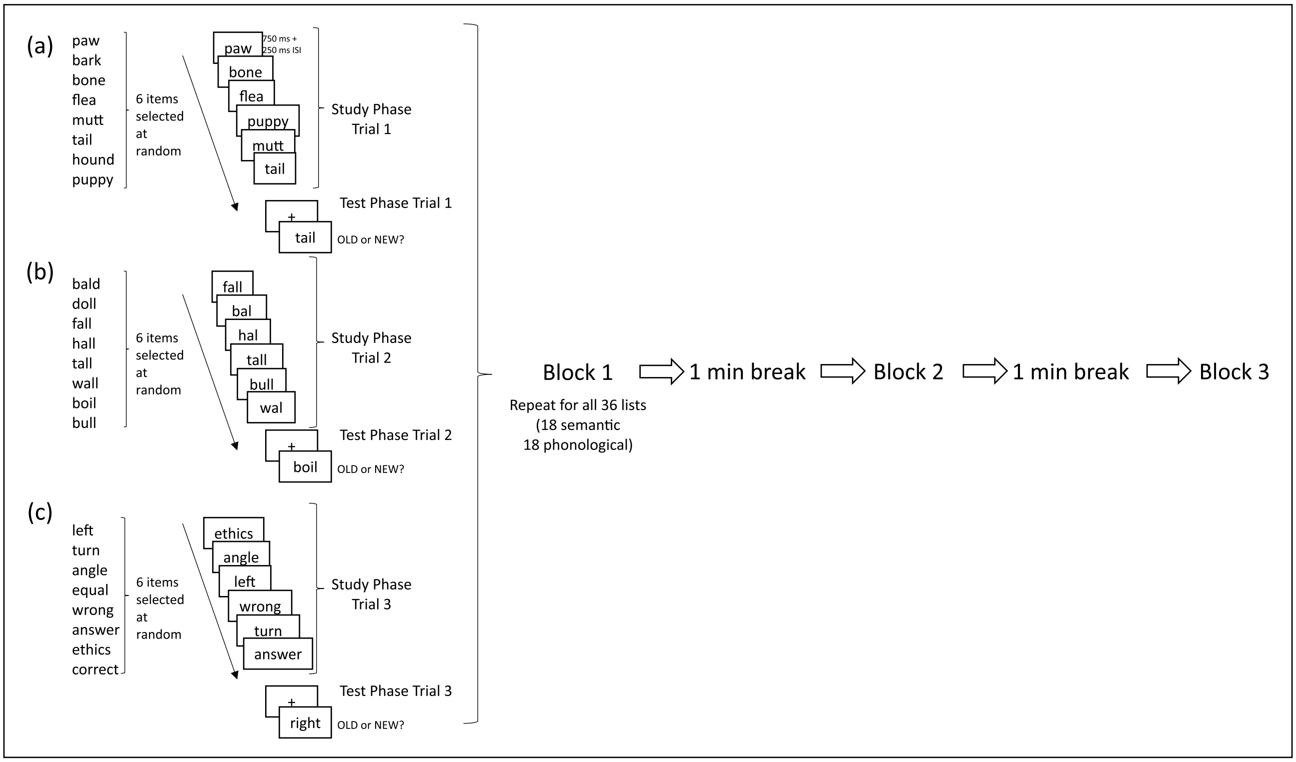

Figure 1. Schematic of experimental procedure in Experiment 1. From each list of eight associates, six items were randomly selected to create a study set. Each item in the set was presented for 750 ms with a 250-ms ISI. (a) An example of a studied probe trial from a semantic list, in which the probe was selected at random from the study set. (b) An example of a non-studied probe trial from a phonological list, in which the probe was selected at random from the two list items that were not included in the set. (c) An example of a CI probe trial from a semantic list, in which the CI related to the list was presented as a test probe. A set corresponding to each of the 36 lists was presented in each of three blocks, with a different probe type in each block. In Experiment 2, each list of associates included 12 items and items were selected from both the semantic and phonological lists to create the seven-item hybrid lists (see the "Method" section for further details on list composition of hybrid lists). The other parameters were the same.

probe was the CI, the studied probe was randomly selected from the studied list, and the non-studied probe was one of the two remaining words in the same associative pool that had not been studied, selected at random. The non-studied probe was included to assess the degree to which errors were sensitive to the strength of the association between list items and non-studied probes. Relative to the CI, which is associated with all items in the list, non-studied probes are likely to be related to a subset of the studied list items and as such are weak associates. Importantly, compared with unrelated probes, they still fit the general theme of the list and thus provide a conservative control condition (cf. Coane et al., 2007; Experiment 2). The studied probe items were the only probe type ever presented in the study lists.

The 108 lists were divided into three blocks of 36 sets corresponding to the 36 different CIs, and blocks were separated by a 1-min break. Thus, a list related to each CI was presented once per block such that each list was presented in each block with six items drawn at random from the eight items listed in the Supplementary Material. For each CI, list type was kept constant across three blocks, and the six items presented in the encoding phase were randomly selected from the phonological or semantic pool anew in each block (thus, items could be repeated and serve as probes in one block and part of the study set in another or vice versa). Probe type was counterbalanced across blocks. Thus, across blocks, each list was tested with a CI probe in one block, a studied probe in one block, and a non-studied probe in one block. Probe order was selected randomly. Two scripts were created so that list type was counterbalanced across these two scripts.[3]

After presentation of the final list, participants were debriefed about the purpose of the study, thanked, and compensated.

Results and discussion

For all analyses that included multiple comparisons, Bonferroni corrections were applied, and where the assumption of sphericity was violated, Greenhouse–Geisser degrees of freedom are reported. We report η_p^2 and Cohen's *d* as measures of effect size. Table 1 presents the

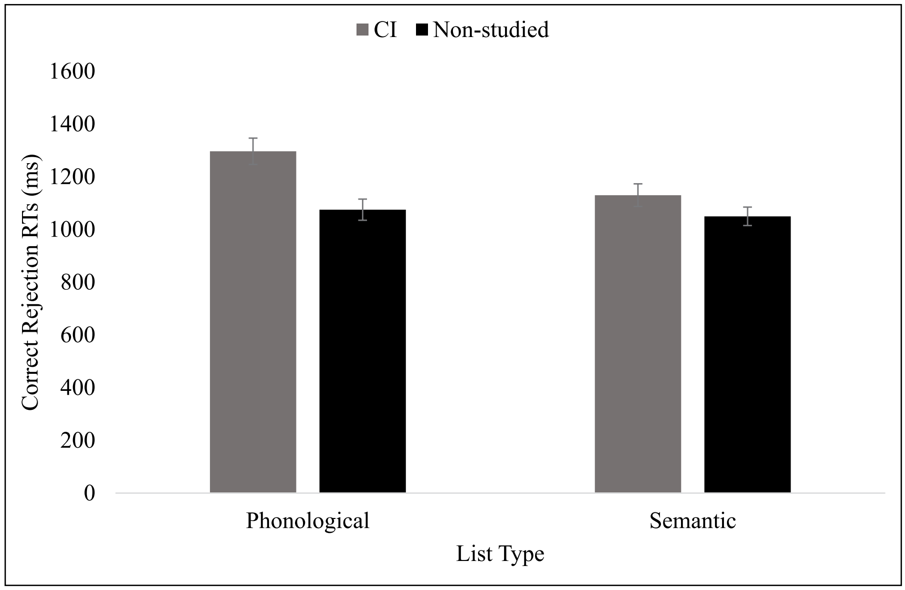

Figure 2. Mean correct rejection RTs as a function of list type and probe type in Experiment 1.
Error bars reflect the standard errors of the means.

Table 1. Mean proportion of old responses as a function of probe type, list type, and block (standard errors in parentheses).

Probe type	Block 1		Block 2		Block 3		Average	
	Semantic lists	Phonological lists	Semantic lists	Phonological lists	Semantic lists	Phonological lists	Semantic lists	Phonological lists
CIs	.16 (.02)	.41 (.02)	.09 (.02)	.34 (.02)	.08 (.02)	.30 (.02)	.11 (.01)	.35 (.02)
Non-studied probes	.08 (.02)	.14 (.02)	.05 (.01)	.09 (.02)	.07 (.02)	.11 (.02)	.07 (.01)	.11 (.01)
Studied probes	.87 (.02)	.79 (.02)	.85 (.02)	.80 (.02)	.78 (.02)	.82 (.02)	.83 (.01)	.80 (.01)

CIs: critical items.

proportion of *old* responses by list type and probe type. Proportions are shown by block because the lists were repeated across the blocks. An examination of the table shows that mean false alarms to CIs were higher for phonological lists than for semantic lists across all three blocks. This result was not seen in the false alarms to non-studied items or in the hits for studied list items. Furthermore, false alarms to CIs were higher than false alarms to non-studied probes in all three blocks for both list types. RTs for *old* and *new* probe responses are shown in Table 2 by list and probe type. Although hit RTs for studied items did not differ across list type, false alarms to CIs and non-studied probes took longer for semantic than for phonological lists. Furthermore, the correct rejection RT comparison for CIs and non-studied items varied according to list type (see Figure 2). These results are supported by statistical tests presented below.

CI versus non-studied probe false alarms. The false alarm rates to CIs and non-studied probes were analysed in a 2×2 repeated measures analysis of variance (ANOVA) with list type (phonological list vs. semantic list) and probe type (CI vs. non-studied probes) as within-subject variables (see Table 1). As noted above, we observed more false alarms for phonological lists ($M = .23$, $SE = .01$) than for semantic lists ($M = .09$, $SE = .01$), $F(1, 100) = 141.34$, mean squared error (MSE) = .014, $p < .001$, $\eta_p^2 = .59$. In addition, there were more false alarms to CIs ($M = .23$, $SE = .01$) than to non-studied probes ($M = .09$, $SE = .01$), $F(1, 100) = 219.84$, $MSE = .009$, $p < .001$, $\eta_p^2 = .69$. The interaction between list type and probe type was also significant, $F(1, 100) = 86.66$, $MSE = .012$, $p < .001$, $\eta_p^2 = .46$.

To further explore the interaction between list type and probe type, we conducted separate paired sample *t* tests. False alarms were higher for CIs than for non-studied probes for both semantic, $t(100) = 5.10$, $SE = .008$, $p < .001$, $d = .52$, and phonological lists, $t(100) = 12.93$, $SE = .018$, $p < .001$, $d = 1.31$, suggesting that false memories for the CIs were found for both semantic lists and phonological lists and that errors were sensitive to the degree of association between list items and non-studied probes. The interaction was driven by the larger difference in false alarm rates between CIs and non-studied probes in phonological lists ($M_{difference} = .24$) than in semantic lists ($M_{difference} = .04$).

Table 2. Mean RTs (ms) to probe responses across list types and probe types in Experiment 1.

List type	Probe type		
	CIs	Non-studied probes	Studied probes
Phonological lists			
M	1,195.04	1,099.28	1,144.58
SD	504.17	491.85	497.05
Semantic lists			
M	1,436.51	1,562.18	1,177.52
SD	1,025.11	1,287.85	601.33

CIs: critical items; RT: reaction times; *SD*: standard deviation.
$N=32$ for CI and non-studied probe false alarm RTs and $N=73$ for hit RTs.

Given the design of the experiment, lists were repeated across the three blocks. This was done to increase the number of observations in each condition. However, as a consequence, participants were re-exposed to list items across blocks. Thus, by the third block, associates to each CI had been presented twice before. This might result in source confusion errors and increase the false alarm rate to CIs due to repeated activation across blocks. Alternatively, the repeated exposure to the lists might improve source monitoring and decrease false alarms to CIs. In addition, non-studied probes might have been repeatedly studied and/or tested in previous blocks.

To address these concerns, we examined false alarm rates including test block as a factor (see means in Table 1). Overall, false alarms decreased over the course of the three blocks from .20 (standard error of the mean [*SEM*] = .01) in Block 1 to .14 (*SEM* = .01) in Block 2 and .14 (*SEM* = .01) in Block 3, $F(2, 200) = 19.0$, $MSE = .02$, $p < .001$, $\eta_p^2 = .16$. Both false alarms to CIs and false alarms to non-studied probes decreased as a function of block, although the decrease was more marked for CIs, especially from Block 1 to Block 2, as evidenced by a significant interaction between probe type and block, $F(2, 200) = 8.26$, $MSE = .019$, $p < .001$, $\eta_p^2 = .08$. Importantly, false alarms to CIs exceeded those to non-studied probes in all three blocks, all $t\text{s} > 6.5$, $SE\text{s} \leqslant .015$, $p\text{s} < .001$, $d\text{s} > .65$, indicating that the differential false alarm rates to CIs and non-studied associates were not due to experimental artefacts. Block did not interact with list type and the three-way interaction was not reliable, both $F\text{s} < 1.0$, $p\text{s} > .42$, indicating that the semantic and phonological similarity effects driving false recognition were similar over the course of the experimental session. Overall, the decrease in false alarms over the course of the experiment does suggest that repeated exposure to list items might have increased participants' monitoring or ability to recognise and reject the CI. However, even after three blocks, false alarms to CIs still exceeded those to non-studied probes.

In sum, consistent with Coane et al. (2007), false alarms to CIs exceeded false alarms to weakly associated non-studied items. Critically, and consistent with memory models that assume differential involvement of phonological and semantic codes in STM and LTM, the false alarm rate for phonological lists was significantly larger than that for semantic lists. The robust effects of list type and probe type emerged even in the first block, indicating that even in the absence of recently activated traces in long-term memory, short-term false memories vary as a function of list type.

CI versus non-studied probe RTs. Because correct rejections reflect situations in which activation and monitoring processes are working in opposition with the result of successful monitoring, we further investigated correct rejection RTs using a separate 2×2 repeated measures ANOVA (see Figure 2).[4] This analysis revealed that correct rejections were slower for phonological lists ($M = 1{,}185.97$, $SE = 43.32$) than for semantic lists ($M = 1{,}089.96$, $SE = 37.57$), $F(1, 99) = 31.65$, $MSE = 29{,}112$, $p < .001$, $\eta_p^2 = .24$. Correct rejections of CIs ($M = 1{,}213.36$, $SE = 44.45$) were slower than correct rejections of non-studied probes ($M = 1{,}062.58$, $SE = 36.28$), $F(1, 99) = 76.01$, $MSE = 29{,}913$, $p < .001$, $\eta_p^2 = .43$. There was a significant interaction between list type and probe type, $F(1, 99) = 15.17$, $p < .001$, $\eta_p^2 = .13$.

Follow-up analyses demonstrated that RTs to CI were slower than RTs to non-studied probes for both phonological, $t(100) = 8.29$, $SE = 26.76$, $p < .001$, $d = .81$, and semantic lists, $t(99) = 3.50$, $SE = 22.93$, $p = .001$, $d = .35$. The interaction was driven by the larger difference between RTs to CI and RTs to non-studied probes in phonological lists ($M_{\text{difference}} = 221.91$) than in semantic lists ($M_{\text{difference}} = 80.15$).

We were unable to perform the analyses on RTs as a function of block because of the small number of valid observations in each condition, which rendered RT data unreliable.

The slower correct rejection latencies to CIs compared with non-studied probes is consistent with differential activation of the two types of probes: CIs are activated by all list items, whereas non-studied probes might only be related to a subset of list items and thus are less activated overall. The higher activation levels require different levels of engagement of cognitive processes to override the strong familiarity signal and reject the CIs.

We also examined participants' false alarm RTs in a similar ANOVA (see Table 2). Not all participants had data in all cells; therefore, the analysis only included data from 39 participants. This analysis only showed a marginal main effect for list type, $F(1, 38) = 3.22$, $MSE = 507{,}829$, $p = .081$, $\eta_p^2 = .08$, with slightly slower RTs for semantic lists ($M = 1{,}506.68$, $SE = 145.64$) than for phonological lists ($M = 1{,}301.88$, $SE = 99.07$). No other effects were significant, both $ps > .40$.

CI versus studied probe accuracy. Finally, we compared the false alarm rates to CIs with the hit rates to studied probes (Table 1) as well as the corresponding RTs (see Table 2). *Old* responses were more frequent for phonological lists ($M = .58$, $SE = .01$) than for semantic lists ($M = .47$, $SE = .01$), $F(1, 100) = 78.20$, $MSE = .015$, $p < .001$, $\eta_p^2 = .44$., with more *old* responses made to studied probes ($M = .82$, $SE = .01$) than to CIs ($M = .23$, $SE = .01$), $F(1, 100) = 854.57$, $MSE = .041$, $p < .001$, $\eta_p^2 = .90$. There was a significant interaction between list type and probe type, $F(1, 100) = 128.28$, $MSE = .014$, $p < .001$, $\eta_p^2 = .56$. The interaction was driven by the fact that *old* responses to studied probes were slightly higher for semantic than for phonological lists, $t(100) = -2.25$, $SE = .013$, $p = .027$, $d = .22$, whereas the opposite effect was observed for CIs, as reported above. This result suggests that the higher false alarm rate for phonological than for semantic lists was not due to a greater tendency for participants to respond *old* to phonological items overall.

To examine whether hit rates changed over the course of the experimental task, we conducted a secondary analysis including block and list type as factors. Overall, hit rates did not change over the course of the experiment, $F(1, 200) = 1.78$, $MSE = .029$, $p = .17$, $\eta_p^2 = .02$. This analysis indicated that in Blocks 1 and 2, hit rates to semantic lists exceeded hit rates to phonological lists (both $ps < .03$); however, the opposite occurred in Block 3 ($p = .016$), as reflected by a significant interaction, $F(2, 200) = 11.95$, $MSE = .018$, $p < .001$, $\eta_p^2 = .12$. Taken in concert with the false alarm data, this suggests that the effect of list type on false recognition occurred regardless of relative veridical recognition performance. In other words, the robust increase in phonological errors relative to semantic errors occurred when hit rates were affected in the same way (i.e., more hits to phonological lists than to semantic lists) or in the opposite direction (i.e., more semantic than phonological hits).

CI versus studied probe RTs. To investigate the effects of list type and probe type on RTs, we again conducted a 2×2 ANOVA. Data from 73 participants were included in this analysis, which revealed faster RTs to probes after phonological lists ($M = 1{,}205.90$, $SE = 50.68$) than those after semantic lists ($M = 1{,}310.17$, $SE = 71.53$), $F(1, 72) = 6.20$, $MSE = 127{,}937$, $p = .015$, $\eta_p^2 = .08$. RTs to CIs ($M = 1{,}355.02$, $SE = 67.98$) were slower than those to studied probes ($M = 1{,}161.05$, $SE = 62.45$), $F(1, 72) = 10.99$, $MSE = 249{,}864$, $p = .001$, $\eta_p^2 = .13$. The interaction between list type and probe type approached significance, $F(1, 100) = 3.27$, $MSE = 113{,}436$, $p = .075$, $\eta_p^2 = .04$. Once again, RT analyses by block were not conducted because of the limited number of observations, especially for false alarms to semantic CIs.

In sum, false alarms to CIs were significantly lower than hit rates, suggesting that participants were able to discriminate between studied and non-studied items. Furthermore, RTs also discriminated *old* responses: False alarms, when they occurred, were slower than hits (see Jou et al., 2004).

Experiment 2

Experiment 2 extended the results of Experiment 1 in a comparison of pure semantic and phonological lists with hybrid semantic/phonological lists of varying composition to examine the separate contributions of semantic and phonological relations to lure false memories at short-term delays. Finley et al. (2017) showed that semantic and phonological list items make separate contributions to long-term false memories that, when combined, create more false memories than pure semantic or phonological lists. Their results showed increasing false memory rates as the lists became more balanced between the two types of items up to a particular level, where activation appeared to plateau once lists contained a few items of the minority feature. We tested whether this result holds for short-term recognition, with hybrid lists of varying composition, ranging from pure semantic and phonological lists of seven items to lists of five items of one type of association and two items of the other type (i.e., two sematic and five phonological items or five semantic and two phonological items). We expected to find the hyper-additivity seen for hybrid list false memories in the Watson et al. (2003) and Finley et al. (2017) studies. However, based on the phonological prominence of short-term/working memory coding (and the results of Experiment 1 showing more false memories for phonological than for semantic lists), we did not expect to replicate the symmetrical contributions of semantic and phonological items seen in Finley et al.'s results for long-term memory. Instead, at short-term delays, phonological items in the list should contribute more than semantic items to the creation of CI false memories and thus yield an asymmetric hyper-additivity pattern. This result would further support the contribution of working memory mechanisms to short-term false memory for phonologically related lists and the dominance of phonological relations for short-term activation effects in CI false memories. These results also highlight the separate contributions of meaning-based and phonological aspects of encoded information to short-term retrieval.

Method

Participants. One hundred and four participants from Colby College ($N=41$) and Illinois State University ($N=63$) participated in this experiment in exchange for course credit. None of the participants had taken part in Experiment 1. The experimental protocol was approved by the IRB at Colby College and at Illinois State University.

Design. The design for Experiment 2 was similar to that of Experiment 1, except for the list type factor. Experiment 2 compared six types of lists: lists of pure semantic associates and phonological associates as in Experiment 1 and four types of hybrid lists that included a mixture of semantic and phonological associates in varying proportions. Each list contained seven associates to the CI. Hybrid lists contained either one semantic or one phonological associate and six associates of the other type (S1P6, S6P1 lists) or two semantic or two phonological associates and five associates of the other type (S2P5, S5P2 lists). List items were drawn at random from a set of 12 items associated with the CI (see the Online Supplementary Material). As in Experiment 1, lists associated with each CI were presented three times across blocks of trials, with a list for each CI presented once per block.

Materials. The 36 CIs and their semantic and phonological associates from the first experiment were used. We added four semantic and four phonological associates to the word list pool for each CI to reduce the likelihood of item repetition across blocks. In Experiment 2, study sets were seven items long instead of six to allow for a more extreme manipulation of list type. For each list presentation, items were chosen from the semantic and phonological word pools according to the ratio of items of each association type for that list. Six set types were created for each DRM list, including pure phonological sets, pure semantic sets, and four hybrid sets. Hybrid sets contained a mixture of phonological and semantic associates. The S0P7 and S7P0 pure sets, where P and S refer to the number of phonological and semantic associates in each set, respectively, consisted of word associates exclusively from the phonological associate pool or the semantic associate pool. The four hybrid sets for each list consisted of associates that were randomly pulled from both of the associate pools in differing proportions. The first of these hybrid sets consisted of one semantic associate and six phonological associates (S1P6), the second consisted of two semantic associates and five phonological associates (S2P5), the third consisted of five semantic associates and two phonological associates (S5P2), and the last consisted of six semantic associates and one phonological associate (S6P1). List length was kept constant across the list types. List types S4P3 and S3P4 were not included because the hybrid effect is most prominent when only one or two different associates are added to the list (Finley et al., 2017).

Six new practice trials were also created where six lists corresponding to six different sets were presented. These lists were not used in the experimental trials. Set type was counterbalanced across six different experimental scripts such that each CI corresponded to a different set type in each script.

Procedure. The procedure was very similar to the first experiment, where participants studied and were tested on 108 sets divided into three blocks. A single recognition probe was presented after the fixation for each seven-item list. The probe selection process occurred as follows: For the non-studied probe, if the majority of the studied list items were semantic or phonological associates, the non-studied probe was chosen from the remaining semantic or phonological associate pool, respectively. This was done to reduce the influence of pop-out effects so that participants could not reject the probe on the basis of the ratio of semantic to phonological studied words. Probe type was again counterbalanced across the three blocks such that each list corresponded to a different probe type across three blocks. Likewise, set types were kept constant across blocks, but the actual studied lists were regenerated in each block.[5]

After the test, participants were thanked, thoroughly debriefed, and compensated.

Results and discussion

The mean proportion of responses for each probe type by list type is shown in Table 3. Although hits rates did not vary much across lists, false alarm rates to CIs varied across the hybrid lists. Specifically, for lists with a majority of phonological associates (S0P7, S1P6, and S2P5), false alarms were higher than those for lists with a majority of semantic associates (S7P0, S6P1, and S5P2). Furthermore, within the lists with a majority of semantic items (S6P1 and S5P2), those with some phonological associates in the lists resulted in more false alarms than the pure semantic list (S7P0). Conversely, false alarm rates were similar across lists with a majority of phonological associates (S0P7, S1P6, and S2P5). Thus, we obtained an asymmetrical hyper-additive effect in short-term memory. Figure 3 shows this pattern in a comparison of false alarms for CIs and non-studied probes across list types. RTs seen in Figure 4 for correct rejections show that overall, correct responses to CIs were slower than correct responses to non-studied items. Statistical analyses reported below support these results.

CI versus non-studied probe false alarms. To assess whether false alarm rates differed as a function of list type, we conducted a 6×2 repeated measures ANOVA with list type (S0P7, S1P6, S2P5, S5P2, S6P1, S0P7) and probe type (CI, non-studied probes) as within-subject variables (see Figure 3). The false alarm rate to CIs ($M=.38$, $SE=.02$)

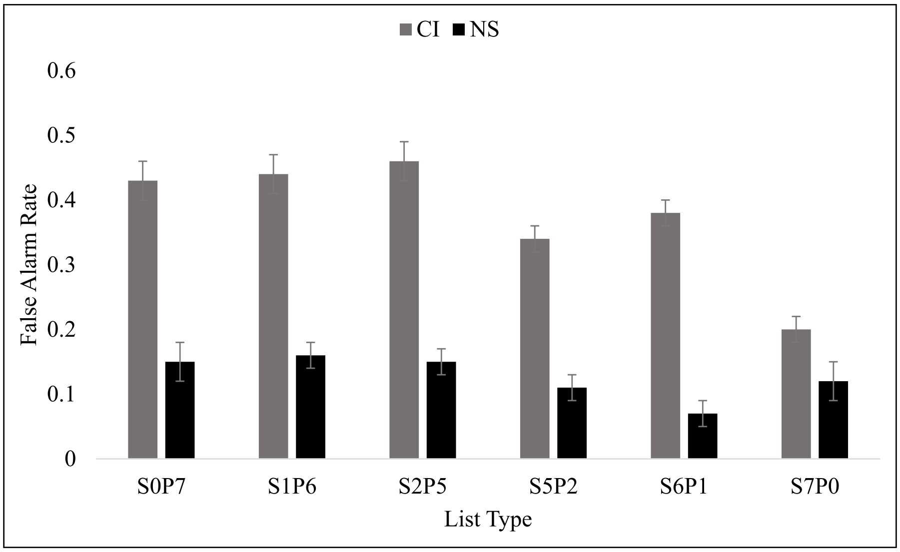

Figure 3. Mean false alarm rate as a function of list type and probe type in Experiment 2.
Error bars reflect the standard errors of the means.

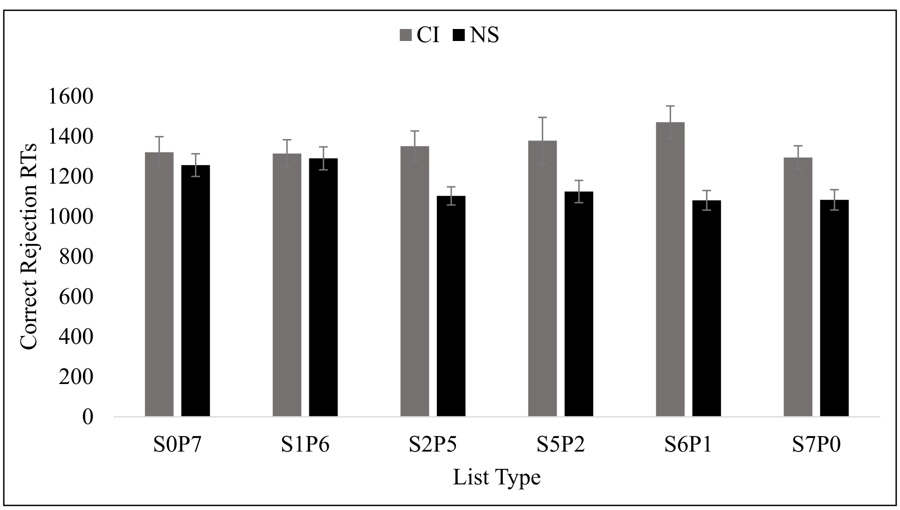

Figure 4. Mean correct rejection RTs as a function of list type and probe type in Experiment 2.
Error bars reflect the standard error of the means.

was higher than that to non-studied probes (M=.12, SE=.02), $F(1, 93)=281.31$, MSE=.063, $p<.001$, $\eta_p^2=.75$. False alarms differed as a function of list type, $F(4.20, 390.71)=17.64$, MSE=.042, $p<.001$, $\eta_p^2=.16$. Follow-up comparisons suggested that false alarm rates were, in general, significantly higher for pure phonological lists (i.e., S0P7) and lists with a majority of phonological associates (i.e., S1P6 and S2P5) than for pure semantic lists (i.e., S7P0) and lists with a majority of semantic associates (i.e., S5P2 and S6P1; all ps < .050, except for the comparison between pure phonological lists and S6P1, which was not significant, p=.12). Furthermore, no significant differences in false alarm rates were observed among phonologically dominant list types (i.e., S0P7, S1P6, and S2P5; all ps > .99). By way of contrast, even though S5P2 and S6P1 lists did not differ significantly from each other in terms of false alarms, $p>.99$, they both elicited higher false alarm rates than pure semantic lists did, both ps < .01.

The main effects were qualified by a significant interaction between list type and probe type, $F(3.72, 346.04)=11.99$, MSE=.031, $p<.001$, $\eta_p^2=.11$. To further explore this interaction, we conducted separate analyses to compare the effects of list type separately for each probe type (see Table 3 for means). The analysis showed that false alarms to CIs were lower for S7P0 than any other list type, all ps < .001. In addition, false alarms to CIs were significantly lower for S5P2 than for both S1P6 and S2P5,

Table 3. Mean proportion old responses for each probe type by list type in Experiment 2.

	List type					
	S0P7	S1P6	S2P5	S5P2	S6P1	S7P0
Studied probes						
M	.75	.73	.78	.81	.78	.76
SD	.17	.22	.21	.21	.19	.17
CI probes						
M	.43	.44	.47	.34	.38	.20
SD	.27	.25	.24	.23	.21	.21
Non-studied probes						
M	.15	.15	.14	.10	.06	.11
SD	.26	.20	.17	.19	.14	.23

CIs: critical items; *SD*: standard deviation; S: number of semantic associates; P: number of phonological associates.

$ps \leq .025$. No other comparisons were significant, all $ps > .05$. Importantly, no significant differences were found among the false alarms to S0P7, S1P6, and S2P5. For non-studied probes, the analysis found that the false alarm rates were significantly lower for S6P1 than for S0P7 ($M_{difference} = .08$), S1P6 ($M_{difference} = .09$), and S2P5 ($M_{difference} = .08$), all $ps < .050$. Other comparisons did not reach statistical significance, all $ps > .20$. Together, these results show that adding phonological associates to pure semantic lists significantly increased the false alarm rates to CIs, but adding semantic associates to pure phonological lists did not achieve comparable effects (cf. Finley et al., 2017; Watson et al., 2003).

We do not report false alarm RTs because only three participants contributed data to all conditions. We also do not report the analyses by block because, given the greater number of conditions, each participant only contributed a maximum of two data points to each cell, thus yielding insufficient observations.

CI versus non-studied probe correct rejection response times. To examine the effects of list type and probe type on the RTs to correct rejection trials (see Figure 4), a 6 × 2 repeated measures ANOVA was conducted with both variables as within-subject factors. Rejecting a CI ($M = 1{,}354.90$, $SE = 63.48$) took longer than rejecting a non-studied probe ($M = 1{,}156.28$, $SE = 46.64$), $F(1, 78) = 47.19$, $MSE = 198{,}141$, $p < .001$, $\eta_p^2 = .38$. The main effect for list type was not significant, $p > .10$, but there was a significant interaction between list type and probe type, $F(3.07, 239.49) = 4.44$, $MSE = 263{,}446$, $p = .004$, $\eta_p^2 = .05$.

To further explore the interaction, we conducted separate *t* tests to compare the effect of probe type on each level of list type, because our primary interest in this analysis was to examine whether opposing effects of activation and monitoring were similar across list type. RTs associated with correctly rejecting CIs were significantly slower than those associated with correctly rejecting non-studied probes when the list types were S2P5, S5P2, S6P1, and S7P0, all $ps < .02$, all $SEs \leq 81.4$. The comparisons did not reach significance for the other two list types (pure phonological lists and S1P6), both $ps > .20$, but numerical trends for these two list types were still consistent with the other list types, showing a slower RT to rejecting CIs than to rejecting non-studied probes. However, this lack of significant difference in the pure phonological and S1P6 lists contrasts with the significant difference between CI and non-studied probe correct rejection RTs found in Experiment 1 for the pure phonological lists. This difference is likely due to the fact that, in Experiment 2, there were fewer observations in each cell of the design, rendering RT data more variable. Nonetheless, participants were generally slower at rejecting CIs than rejecting non-studied probes, consistent with the results of Experiment 1.

CI versus studied probe Old responses. Finally, we compared the false alarm rates with CIs and the hit rate to studied probes as well as the corresponding RTs. We conducted a 6 × 2 ANOVA as in the previous analysis (see Table 3). Hit rates to studied probes ($M = .77$, $SE = .01$) were significantly higher than false alarms to CIs ($M = .38$, $SE = .02$), $F(1, 103) = 399.69$, $MSE = .119$, $p < .001$, $\eta_p^2 = .80$. There was a significant main effect for list type, $F(4.51, 464.21) = 15.55$, $MSE = .036$, $p < .001$, $\eta_p^2 = .13$. The follow-up comparisons showed that the proportion of *old* responses was significantly lower for pure semantic lists than for any other list type, all $ps < .001$. Furthermore, S2P5 appeared to elicit a higher proportion of old responses than S5P2, $p = .019$. No other comparison was significant, all $ps > .290$. There was also a significant interaction between list type and probe type, $F(4.38, 451.19) = 15.76$, $MSE = .04$, $p < .001$, $\eta_p^2 = .13$.

To further investigate the interaction, we conducted a one-way ANOVA to compare hit rates across list types. The analysis comparing false alarm rates across list types was reported above. The analysis for hit rates revealed a significant effect for list type, $F(4.40, 453.03) = 2.95$, $MSE = .032$, $p = .017$, $\eta_p^2 = .03$. The hit rate was significantly higher only for S1P6 lists than for S5P2 lists, $p = .042$ (for all other comparisons, $ps > .35$). Thus, overall, hit rates were relatively stable across list types.

CI versus studied probe Old response times. We next examined the effects of list type and probe type on hit and false alarm RTs, using the same ANOVA. RTs to false alarms ($M = 1{,}185.58$, $SE = 59.98$) were significantly slower than RTs to hits ($M = 1{,}085.55$, $SE = 48.26$), $F(1, 50) = 12.16$, $MSE = 1{,}530{,}798$, $p = .001$, $\eta_p^2 = .20$. The effect of list type was also significant, $F(4.07, 203.41) = 4.48$, $MSE = 130{,}573$, $p = .002$, $\eta_p^2 = .08$. Follow-up tests suggested that RTs were faster for S0P7 lists than for S6P1 lists, $p = .028$, and S2P5 lists were faster than S7P0 lists, $p = .024$. No other comparisons were significant, all $ps > .06$. These main effects were further qualified by a significant interaction between list type and probe type, $F(3.30, 164.96) = 7.90$, $MSE = 214{,}132$, $p < .001$, $\eta_p^2 = .14$.

To further explore this interaction, we conducted six t tests to compare the effect of probe type at each level of list type. RTs to CIs were slower than those to studied probes when the list types were S0P7, S7P0, and S6P1, all $ps \leq .01$, $SEs \leq 89.1$. No other comparisons were significant, all $ps > .050$. Thus, it seemed that participants were generally slower in identifying CIs as *old* than at identifying studied probes as *old*, but this effect was reduced as the lists contained a more balanced mixture of the two types of items (i.e., more semantic/phonological associates were added to pure phonological or semantic lists).

General discussion

The current study produced two novel findings: (a) phonologically related lists created more short-term false memories than semantically related lists, with a reversed effect of list type on hit rates (i.e., more accurate identification of studied list items for semantic than phonological lists) and (b) semantic/phonological hybrid lists showed asymmetric hyper-additivity in false memory rates, with a stronger contribution of phonological items than semantic items. The former result shows a phonological similarity effect in short-term false memories, and the latter result further suggests that phonological similarity overshadows semantic similarity effects on short-term tests. Both of these results suggest a working memory contribution to short-term false memories for phonological lists because long-term memory contributions should increase false memories for semantic lists. We address the importance of each of these findings to the current theoretical descriptions of short-term false memories below.

Activation–monitoring account

Previous findings of short-term false memories for semantically related CIs (e.g., Atkins & Reuter-Lorenz, 2008; Coane et al., 2007) have been described in terms of the activation–monitoring account of false memories (Roediger et al., 2001). This description of false memories assumes that the CIs are activated by related items presented in the studied lists. At test, this activation can be mistaken for familiarity due to study when source monitoring errors occur. Finley et al. (2017) describe their hyper-additive false memory effects for hybrid lists in this way, with separate activation functions assumed for semantic and phonological associations. Coane et al.'s results for short-term false memories support this account: Higher false alarms for CIs than weakly associated items from semantic lists were found for list lengths of 5 and 7, but not for lists of three items, suggesting that the shorter lists did not contribute enough activation of the CIs to lead participants to endorse them as studied in the memory test. The small memory set in that study likely also made monitoring processes more effective and successful. The results in the current study are also consistent with an activation–monitoring description of false memories. Given the phonological coding dominance of information in short-term/working memory (e.g., Nairne, 1990, 2002), activation of phonological CIs is stronger than that of semantic CIs during short-term tests. The present results of higher false alarm rates to CIs for phonological than for semantic lists support this description. Higher hit rates for semantic than for phonological lists suggest that the false alarm data are not due to a mere bias to respond *old* to phonological list probes (although it is possible that this result is due to higher rates of discrimination for the semantic than for the phonological lists). In addition, the hybrid lists used in Experiment 2 resulted in an asymmetric pattern of CI false alarms—lists with a higher proportion of phonological items did not differ in false alarm rates but lists with a higher proportion of semantic items displayed the hyper-additive pattern found for hybrid lists in long-term memory in past studies (Finley et al., 2017; Watson et al., 2003). Furthermore, as seen in Figure 3, phonologically based hybrid lists produced higher levels of false memory overall compared with semantically based hybrid lists.

The asymmetric function shown in the current results contrasts with the function reported by Finley et al. (2017) for phonological/semantic hybrid lists in standard long-term memory measures. Their lists ranging from pure semantic or phonological lists to hybrid lists with equal numbers of phonological and semantic items resulted in a symmetric hyper-additive function of CI false alarms. However, their participants were given a free recall test after each list. Free recall tests encourage reliance on list item relations more than single-item recognition tests do (Hunt & Einstein, 1981); thus, this could explain the difference between their false alarm function and the one found in the current study for hybrid lists. The greater reliance on item-specific information in the single-item short-term recognition tests in the current study would result in more phonological confusion for hybrid lists with higher proportions of phonological items than those of semantic items. Because item-specific information (in terms of orthography and/or phonology) differs more for semantic list items and between semantic associates and CIs, CIs are more easily rejected with a higher proportion of semantic items in the hybrid lists. In addition, the asymmetry highlights the different levels of reliance on phonological versus semantic information in short- and long-term memory. Importantly, however, the present results do indicate that hyper-additive effects emerge even in the short term, thus providing an important extension and boundary condition to the work by Watson et al. (2003) and Finley et al. Future studies will examine more directly the dissociation in hybrid false memory effects between short- and long-term memory.

One point that merits a brief discussion is the finding that in Experiment 1, hit rates to semantic lists exceeded those to phonological lists in two of the experimental blocks and overall. At first blush this would appear inconsistent with the false alarm data, in that the greater similarity of list items to one another, which increases phonological false alarms, might also be expected to increase hit rates, if participants are relying on similarity to drive memory decisions. However, the semantic advantage in hit rates might reflect a similar process as discussed above, namely, that the semantic associates provide more distinctive orthographic and phonological traces and are thus less confusable. The effects of distinctiveness in long-term memory performance are well established (e.g., Hunt & McDaniel, 1993) and temporal distinctiveness has been proposed as a key factor in short-term memory (e.g., Neath & Brown, 2006; but see Oberauer & Lewandowsky, 2008, for alternative accounts). Such an explanation would be consistent with the role of a monitoring component as proposed by the activation–monitoring account: More distinctive traces would be easier to accept as *old* when studied and reject as *new* when non-studied. As noted in the Introduction, the effects of phonological similarity in short-term memory have been well documented; the present results add to the literature on phonological similarity effects by confirming the increased error rates as a function of overlap in phonological features that discriminated between errors to CIs and to non-studied associates.

Further support for the activation–monitoring account is seen in the RT data for correct rejections. Overall, correct rejections to CIs were slower than to unstudied probes, consistent with the conflict between activation and monitoring expected for CIs. These results are in line with those reported by Coane et al. (2007) and Atkins and Reuter-Lorenz (2008) for correct rejections of CIs from semantic lists. This effect was particularly pronounced in Experiment 1 of the present study with phonological lists, where the difference in RTs between CIs and unstudied probes was larger than for semantic lists, suggesting that phonological activation was especially strong in this short-term memory test. This result replicated in Experiment 2 for most of the hybrid lists and the pure semantic list. It was noticeably absent, however, in the pure phonological and S1P6 lists. This might be due to the contrast across list types in the within-subjects design or to the increased variability in RTs due to the smaller number of observations.

The activation–monitoring description of short- and long-term false memory is consistent with a unitary memory account (e.g., Cowan, 1999; Nairne, 2002). False memories have been found in past studies for semantically related lists across both short- and long-term memory tests (Atkins & Reuter-Lorenz, 2008), further supporting a consistent description of false memory processes across the short and long terms. Flegal and Reuter-Lorenz's, (2014) results of similar phenomenology of CI false alarms in a remember-know design are also consistent with this idea. Thus, the studies to date of short-term false memories provide strong support for consistent short- and long-term memory mechanisms. What does appear to differ is the level of semantic and phonological contributions across the time course.

Fuzzy trace theory

An alternative explanation of false memory effects in long-term memory, known as fuzzy trace theory, suggests that two types of memory traces can be accessed in retrieval—verbatim and gist traces (Brainerd & Reyna, 2002, 2005). Gist and verbatim traces can be accessed for studied items, but gist is the cause of false alarms to and false recall of CIs. Because gist is similar to that for studied items, retrieval of the gist when evaluating a CI as a studied item causes it to be falsely remembered as a list item.

False memories for short- and long-term tests with semantic lists are consistent with the fuzzy trace explanation. However, Flegal and Reuter-Lorenz, (2014) argued that their results showing no level of processing effect for "remember" responses to CIs at short- and long-term delays were inconsistent with fuzzy trace theory, because decay of verbatim traces over time should dissociate the phenomenology of these false memories across delays. They stated that deep encoding at study should strengthen verbatim traces, making it more likely that participants could use this information to reject CIs at the short-term test, reducing the likelihood they would "remember" studying the CIs. But no difference was found in "remember" responses. Fuzzy trace theory, at least in its traditional instantiation where gist is described as being based on shared meaning, also has difficulty explaining the current results with phonological lists. Finley et al. (2017) argued that because false memories are based on gist traces, according to the current conceptualisation of the theory, it cannot account for long-term phonological false memories, because gist is assumed to be based on the semantic features of the list items, not on shared surface features. The increased false memories for phonological lists compared with semantic lists found in the current study is also problematic because verbatim traces should still be available at the short term to reduce false alarms and there is no gist contribution for phonological lists to increase false alarms for these CIs. It is possible that greater confusion of verbatim traces for phonological lists than semantic lists caused the higher phonological false alarm rate in our first experiment. This would be consistent with Olszewska et al.'s (2015) results showing lower CI false alarms for semantic lists when they were presented auditorily than visually. However, it does not explain the hyper-additivity seen in our hybrid lists in Experiment 2. Hybrid lists with some semantic items should make it easier to distinguish verbatim traces due to fewer confusable phonetic traces. Therefore, without further modification of fuzzy trace theory, it cannot easily account for the current results or others' findings of phonological long-term false memories (e.g., Finley et al., 2017; Sommers & Lewis, 1999; Watson et al., 2003). As noted in the Introduction, the concept of phonological gist has been proposed (e.g., Holliday & Weekes, 2006); however, this extension of gist extraction along sources of similarity beyond shared meaning has not been explicitly incorporated into the theoretical framework. Such a modification could be fairly straightforward and would expand the concept of gist to refer to the process by which statistical regularities across items that share specific features are extracted. However, this approach might still render the explanation of the results from hybrid lists problematic, because it would need to assume that multiple gist traces (i.e., semantic and phonological) are extracted and stored in parallel, which seems potentially lacking in parsimony.

Time-based resource sharing model of working memory

Abadie and Camos (2018) applied the time-based resource sharing model of working memory to their short- and long-term false memory results. This model suggests that working memory relies on two mechanisms to hold verbal information—articulatory rehearsal and attentional refreshing. They hypothesised results for their experiments testing semantic false memories for short- and long-term delays in terms of the effects of these processes on gist and verbatim traces as described by fuzzy trace theory. In their first two experiments, Abadie and Camos manipulated the availability of articulatory rehearsal for short- and long-term recognition of semantically related lists. Their results indicated that at short-term tests, false memories to the CIs were only present when working memory rehearsal was prevented, but false memories occurred without rehearsal in the long-term tests. Their third and fourth experiments examined the effects of attentional refreshing and found that refreshing reduces false memories at the long term. They also showed (through model fits) that participants relied more on the gist than on the verbatim traces of the studied items for responses in the short-term tests when rehearsal was blocked. From these results, they concluded that working memory does not contribute to false memories at the short term when rehearsal is allowed because rehearsal strengthens the verbatim traces of list items, making it easier to reject CIs as non-studied items. They further suggested that gist retrieved from long-term memory is the cause of the false memories found at both the short and long terms and that when articulatory rehearsal processes operate at the short term, verbatim traces of the studied items prevent false memories at short-term tests.

The current results do not clearly align with Abadie and Camos' (2018) description of working memory contributions to short-term false memories. Because we did not prevent rehearsal or attentional refreshing, our results for both types of lists should be similar to results from their first experiment: few false alarms for CIs and no difference in false alarms for CIs and weakly related probes. The higher rates of false alarms for CIs from phonological lists (and hybrid lists with a higher proportion of phonological items in Experiment 2) across the current experiments are also difficult to explain based on reliance of gist and verbatim traces as noted above. Confusion of verbatim traces for the phonological lists does not explain the hyper-additivity of the hybrid lists seen in our Experiment 2. Thus, although we did not specifically test the contributions of working memory on false memories in our study, our results are not consistent with Abadie and Camos' conclusion that working memory processes do not play a role in false memories at the short term. Rather, our results are more in line with those reported by Olszewska et al. (2015) who showed that presentation and test modality of semantic lists dissociated short- and long-term false memories. At short-term tests in their first experiment, they showed fewer false alarms to semantic CIs for the auditory than for the visual presentation modality. Our phonological lists had the opposite effect—we found more false alarms to CIs for the phonological lists than for the sematic lists, due to the verbal coding dominance present in working memory.

Summary

The current experiments generalised findings of short-term false memories in semantically related DRM lists to phonologically related lists. Based on our results showing higher rates of false memories for phonological than for semantic lists and especially hyper-additive effects of hybrid lists, we favour an activation–monitoring description of short-term false memories. Future work should explore the specific contributions of articulatory rehearsal processes to activation of CIs for phonological and semantic lists during short-term memory tests.

Acknowledgements

We thank Jeffrey Kahn for consulting on statistical analyses. We also thank Andrew Parker, Henry Roediger, and an anonymous reviewer for helpful comments on an earlier version of the manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by funding from a James McDonnell Foundation *Understanding Human Cognition* Grant (#220020426) awarded to JHC. The funding agency had no input on the study design, data analysis, or writing.

Supplementary material

The Supplementary Material is available at: qjep.sagepub.com

Notes

1. Although there are theoretical reasons to distinguish between conceptual/semantic and lexical/associative relations in semantic memory (see Hutchison, 2003; Lucas, 2000, for reviews), in the context of the present study, we use the term "semantically related" or "semantic lists" to contrast with relations driven by orthographic and phonological similarity, as described below.
2. Whereas rehearsal maintenance has been a key aspect of working memory theories for decades, attentional refreshing is a relatively recent addition to working memory models that uses attention to reactivate stored information (Camos, 2015). Both processes can be used to maintain information in short-term memory. However, disruption of either process (using articulatory suppression or a divided attention task) will reduce memory performance.
3. Due to an error in the programming code, the number of probe types tested in each block was not exactly one third; over the course of the experiment, however, the relative proportion of trials with each probe type was balanced. Furthermore, the counterbalancing error was not systematic and as such was distributed across participants.
4. The difference in degrees of freedom reflects the fact that not all participants had correct rejections; thus, their data are not included in this analysis.
5. Due to a programming error, the randomisation between non-studied and studied probes in the pure lists (S0P7 and S7P0) was unevenly distributed across participants. For all hybrid lists, the counterbalancing was exact. Importantly, even for pure lists, the overall pattern of results is consistent with those obtained in Experiment 1, suggesting that the variability in relative proportions of studied to non-studied probes was not a major factor in the results.

ORCID iD

Dawn M McBride https://orcid.org/0000-0002-5357-2913

References

Abadie, M., & Camos, V. (2018). False memory at short and long term. *Journal of Experimental Psychology: General*. Advance online publication. doi:10.1037/xge0000526.

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition*, *36*, 74–81. doi:10.3758/MC.36.1.74

Baddeley, A. D. (1964). Semantic and acoustic similarity in short-term memory. *Nature*, *204*, 1116–1117. doi:10.1038/2041116a0

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy trace theory and false memories: New frontiers. *Journal of Experimental Child Psychology*, *71*, 194–209. doi:10.1006/jecp.1998.2472

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164–169. doi:10.1111/1467-8721.00192

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York, NY: Oxford Press.

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition*, *51*, 251–261. doi:10.1016/S0278-2626(03)00030-7

Camos, V. (2015). Storing verbal information in working memory. *Current Directions in Psychological Science*, *24*, 440–445.

Coane, J. H., & McBride, D. M. (2006). The role of test structure in creating false memories. *Memory & Cognition*, *34*, 1026–2036. doi:10.3758/BF03193249

Coane, J. H., McBride, D. M., Raulerson, III., & Jordan, J. S. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*, 62–70. doi:10.1027/1618-3169.54.1.62

Coane, J. H., McBride, D. M., Termonen, M. L., & Cutting, J. C. (2016). Categorical and associative relations increase false memory relative to purely associative relations. *Memory & Cognition*, *44*, 37–49. doi:10.3758/s13421-015-0543-1

Conrad, R. (1963). Acoustic confusions and memory span for words. *Nature*, *197*, 1029–1030. doi:10.1038/1971029a0

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York, NY: Cambridge University Press.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22. doi:10.1037/h0046671

Dodd, M. D., & MacLeod, C. M. (2004). False recognition without intentional learning. *Psychonomic Bulletin & Review*, *11*, 137–142. doi:10.3758/BF03206473

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Finley, J. R., Sungkhasettee, V. W., Roediger, H. L., III., & Balota, D. A. (2017). Relative contributions of semantic and phonological associates to over-additive false recall in hybrid DRM lists. *Journal of Memory and Language*, *93*, 154–168. doi:10.1016/j.jml.2016.07.006

Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *36*, 1331–1338. doi:10.1037/a0019903

Flegal, K. E., & Reuter-Lorenz, P. A. (2014). Get the gist? The effects of processing depth on false recognition in short-term and long-term memory. *Memory & Cognition*, *42*, 701–711.

Frost, P., Ingraham, M., & Wilson, B. (2002). Why misinformation is more likely to be recognised over time: A source monitoring account. *Memory*, *10*, 179–185. doi:10.1080/09658210143000317

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848. doi:10.3758/MC.38.7.833

Holliday, R. E., & Weekes, B. S. (2006). Dissociated developmental trajectories for semantic and phonological false memories. *Memory*, *14*, 624–636. doi:10.1080/09658210600736525

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*, 497–514. doi:10.1016/S0022-5371(81)90138-9

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421–445. doi:10.1006/jmla.1993.1023

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*, 785–813. doi:10.3758/BF03196544

Jou, J., Matus, Y. E., Aldridge, J. W., Rogers, D. M., & Zimmerman, R. (2004). How similar is false recognition to veridical recognition objectively and subjectively? *Memory & Cognition*, *32*, 824–840.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. doi:10.1037/0033-2909.114.1.3

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*, 618–630. doi:10.3758/BF03212999

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36. https://doi.org/10.1097/00003446-199802000-00001

Murdock, B. B., Jr. (1961). The retention of individual items. *Journal of Experimental Psychology*, *62*, 618–625. doi:10.1037/h0043657

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, *21*, 263–276. doi:10.1037/h0082978

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.

Nairne, J. S., & Neath, I. (2013). Sensory and working memory. In A. F. Healy & R. W. Proctor (Eds.), *Comprehensive handbook of psychology: Experimental Psychology* (2nd ed., *Vol.4*, pp. 419–445). New York, NY: John Wiley & Sons.

Neath, I., & Brown, G. D. A. (2006). SIMPLE: Further applications of a local distinctiveness model of memory. *Psychology of Learning and Motivation*, *46*, 201–243. doi:10.1016/S0079-7421(06)46006-0

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, *143*, 992–1009. doi:10.1037/bul0000108

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, *115*, 544–576. doi:10.1037/0033-295X.115.3.544

Olszewska, J. M., Reuter-Lorenz, P. A., Munier, E., & Bendler, S. A. (2015). Misremembering what you see or hear: Dissociable effects of modality on short- and long-term false recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *41*, 1316–1325. doi:10.1037/xlm0000115

Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin*, *82*, 68–84. doi:10.1037/h0076166

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*, 398–422. doi:10.3758/BF03202613

Pérez-Mata, M. N., Don Read, J., & Diges, M. (2002). Effects of divided attention and word concreteness on correct recall and false memory reports. *Memory*, *10*, 161–177. doi:0.1080/09658210143000308

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, *5*, 1–9. doi:10.1016/j.jarmac.2015.12.003

Roediger, H. L., III., Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger III, 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. doi:10.1037/10394-006.

Roediger, H. L., III., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 803–814. doi:10.1037/0278-7393.21.4.803

Roediger, H. L., III., 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. doi:10.3758/BF03196177

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Sommers, M. S., & Lewis, B. R. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language*, *40*, 83–108. doi:10.1006/jmla.1998.2614

Sugrue, K., Strange, D., & Hayne, H. (2009). False memories in the DRM paradigm: Age-related differences in lure activation and source monitoring. *Experimental Psychology*, *56*, 354–360. doi:10.1027/1618-3169.56.5.354

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, *29*, 424–432. doi:10.3758/BF03196393

Watson, J. M., Balota, D. A., & Roediger, H. L., III. (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory and Language*, *49*, 95–118. doi:10.1016/S0749-596X(03)00019-6

Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *The American Journal of Psychology*, *78*, 567–574. doi:10.2307/1420917