



The influence of semantic context on false memories

Durna Alakbarova¹ · Jason L. Hicks² · B. Hunter Ball¹

Accepted: 20 April 2021
© The Psychonomic Society, Inc. 2021

Abstract

Although false memories have largely been examined with the Deese-Roediger-McDermott (DRM) paradigm, little research has focused on the semantic context in which associates are encoded. Across three experiments, we varied semantic context during a sentence-processing task with DRM associates embedded within sentences. More meaningful sentences resulted in greater memory errors (Experiment 1). Furthermore, providing contextual information to discriminate old from new items did not reduce false alarms relative to encoding words in isolation when sentences converged on the meaning of the critical lure (Experiment 2), and actually increased memory errors (Experiment 3). These results suggest that semantic context that allows for meaningful relational processing of items within-lists and that converge on the semantic meaning of the critical lure increases the likelihood that the list theme is identified, resulting in more errors at test.

Keywords False memory · Recognition memory · DRM · Context

Introduction

Examination of the factors involved in the creation of false memories provides a fruitful method of investigating the underlying mechanisms involved in the organization of human memory. Memory errors have been largely studied using a variety of methods and have been found to occur when there is an overlap in semantic, orthographic, or phonological features between old and new items (Gallo, 2013), when there is not a sufficient discrimination in source memory of the items (e.g., Winograd, 1968), when presented with “fake news,” especially if the content aligns with held beliefs (Greene et al., 2020), in instances of unconscious plagiarism (e.g., Marsh et al., 1997), as well as during eyewitness testimony (e.g., Loftus, 1971) and free-recall tests (Unsworth & Brewer, 2010). A method that has been widely used to investigate semantic false memories is the Deese-Roediger-McDermott (DRM) paradigm, whereby after studying a list of related words such as *bed*, *rest*, *tired*, and *dream*, people often erroneously claim that a non-presented critical lure (*sleep*) was originally studied (Deese, 1959; Roediger & McDermott,

1995). The current study was designed to examine false memories in the DRM recognition memory paradigm by varying the semantic context in which associates are encoded during a sentence-processing task.

Central to the underlying mechanisms involved in false memories using the DRM paradigm is the organization of list items. Each item within a list is associated (by frequency of co-occurrence) with a common theme, which can include orthographic, categorical, or conceptual similarities between the items and theme. The typical DRM paradigm presents list items in a blocked fashion according to theme and has participants try to remember these items for a later memory test. Relating items in terms of categorical or conceptual relationships benefits veridical memory performance but may also increase susceptibility to false memory errors (McCabe et al., 2004; Toglia et al., 1999). One theory that accounts for such findings is the activation/monitoring theory (AMT; Roediger et al., 2001). This theory posits that activation of critical lures occurs during processing of list items via spreading activation of conceptual representations within a semantic network (Anderson, 1983; Collins & Loftus, 1975; Gallo, 2010). During encoding, a summation of multiple implicit associative responses produced by the studied associates may internally activate the conceptual representation of the critical lure, thus making it available in memory (Hancock et al., 2003; Underwood, 1965). During retrieval, test probes may serve to reactivate the associative network that subsequently makes the critical lure susceptible to false remembering due to a high degree of overlap between the lure and its activated

✉ B. Hunter Ball
Hunter.Ball@uta.edu

¹ Department of Psychology, University of Texas at Arlington, Arlington, TX 76019, USA

² Department of Psychology, Louisiana State University, Baton Rouge, LA, USA

representation within the associative network (Kimball et al., 2010; Meade et al., 2007). False memories therefore occur due to a reality-monitoring error (Johnson et al., 1993) in which participants mistake internally generated items as actually being perceived (e.g., Hicks & Marsh, 1999, 2001).

An alternative theory that has been proposed to explain the DRM illusion is the fuzzy-trace theory (FTT; Brainerd & Reyna, 2002; Reyna & Brainerd, 1995). This theory suggests that encoding of list items results in two types of memory traces: verbatim and gist. Verbatim traces include specific contextual details from processing surface forms of experienced items, whereas gist traces reflect extracted commonalities among experiences whereby participants mentally construct a gist representation of common features of the conceptual form of items. Because no verbatim traces are present for critical lures, false alarms to critical lures are thought to occur because the lure is highly similar to the gist representation. This produces a strong feeling of familiarity (or in some instances, "phantom recollection"), causing participants to mistakenly call the item "old" (Brainerd et al., 2001). False alarms can be reduced by recollecting verbatim traces from the study episode (e.g., bed), which allows participants to reject the item as having occurred previously.

The primary difference between the two theories is that AMT assumes that associative activation occurs due to statistical co-occurrences of items within a mental lexicon, whereas FTT posits that gist extraction occurs due to emergent semantic properties from the list structure. While there is some evidence to suggest that lure false alarms (e.g., salt) can occur for semantically related items that have no associative relation (e.g., butter; Brainerd et al., 2008), in general these theories make similar predictions for DRM studies. This is by virtue of the DRM list structure, which confounds associative and semantic relations (Brainerd et al., 2020). That is, because items that are associatively related are also typically semantically related, manipulations that influence lure activation typically influence gist processing. Both theories also suggest that recollective processes at retrieval (i.e., source monitoring vs. recollection rejection) can be used to counteract the strong feelings of familiarity produced by the critical lure. For ease of exposition, we therefore use the terminology described by FTT (i.e., gist and recollective retrieval) as it has most commonly been applied to describe previous research related to the current set of studies. However, we return to these theoretical distinctions in the *General discussion* and note how AMT can similarly be explained to the findings.

In addition to the distinction between verbatim and gist traces, Neuschatz et al. (2002) argued that for words, gist can be further broken down into local and global gist. Local gist reflects meaning extracted from items considered in isolation, whereas global gist reflects the extraction of meaningful relations among items (Lampinen et al., 2006; Odegard et al., 2008). Findings that even a single presentation of an

associate (e.g., bed) increases false alarms to critical lures (e.g., sleep) highlights the importance of local gist processing (Underwood, 1965). Semantic orienting tasks and deeper levels of processing also increase veridical and false memories (Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toggia et al., 1999), suggesting that focusing on the meaning of items can increase local gist extraction (Odegard et al., 2008). However, global list structure is important, as it can orient processing towards noticing the relation *among* items within a list. For example, associative lists blocked by theme increase true and false memories relative to random list presentation (Brainerd et al., 2003; McDermott, 1996; Toggia et al., 1999), and intermixing unrelated filler items within blocked lists reduces false memories (Goodwin et al., 2001). Because local gist processing is equated across list format (i.e., same number of semantically related list items), this suggests that random presentation disrupts global gist processing (Goodwin et al., 2001; Lampinen et al., 2006). Instructions that encourage relational processing among items within a list also increase false memories relative to instructions that focus on item-specific features of each list item (McCabe et al., 2004), and distinctive processing at encoding (e.g., unique fonts) has been shown to reduce false memories (Arndt & Reder, 2003). These findings suggest that it is important to not only consider the nature of the studied associate, but also how meaningful relations can be formed within a list based on the subjective organization imposed by the participant during encoding (Gallo, 2013) memories.

Another manipulation to examine the influence of semantic processing on false memories is embedding associates within sentences or text. According to the discourse comprehension literature, in addition to verbatim traces, multiple levels of gist can be extracted from a text (i.e., sentence, text, and situation models; Clark & Clark, 1977; Glucksberg & Danks, 1975; Kintsch et al., 1990). Readers may extract verbatim details of the exact wording of the sentences (e.g., "After work he lay down on the bed. He had a frightening dream"), but also sentence-level gist (e.g., he was tired and scared) and story level-gist (e.g., he was sleeping). This is similar to the distinction between local and global gist in the DRM paradigm, where words within a sentence make up the local meaning of the semantically related associate (e.g., bed) and the list as a whole makes up a more global structure associated with the critical lure (e.g., sleep). Because the sentences in the previous example converge on a common theme, presumably both local and global gist processing should occur. However, sentences that diverge from the theme should disrupt global gist processing (e.g., "She walked along the river *bed*"; "The new car drives like a *dream*"; etc.).

Several studies have used story contexts that converge on a central theme to examine developmental trends in false memory. The typical finding during standard list processing is that younger children (e.g., age 5 years) show fewer false

memories than older children (e.g., age 11 years) or adults. However, when DRM associates are embedded within sentences during story processing, age differences between younger and older children are attenuated (Howe & Wilkinson, 2011; Swannell & Dewhurst, 2013) or even eliminated (Dewhurst et al., 2007). Similarly, children actually show higher false-alarm rates than adults when associates are embedded in stories (Otgaar et al., 2014). Although the exact mechanisms are debated, the general explanation for this pattern of findings is that younger children have not fully developed their mental lexicon, meaning that during standard list processing they are less likely to notice relations among items and extract the global gist. However, story contexts make it easier for these children to identify the overall theme, thereby elevating false alarms towards rates of older children (or adults) with more fully developed lexicons. Interestingly, item-specific processing (e.g., focusing on spelling) reduces false memories more for older children than for younger children (Holliday et al., 2011).

On the opposite end of the developmental spectrum, it has been shown that while younger adults are able to use contextual details to reduce false memories when associates are presented in individual sentences that converge on the meaning of the critical lure relative to standard list processing (e.g., *bed*; Thomas & Sommers, 2005), older adults (> 60 years old) are not. Only when sentences diverge from the meaning of the critical lure are older adults able to reduce false memories relative to list processing in a similar manner to younger adults (for similar results in children see Howe & Wilkinson, 2011). Notably, Thomas and Sommers (2005) found that response latencies for false alarms to critical lures were similar for convergent and divergent conditions for both age groups, suggesting that both lure types were activated. This suggests that younger adults are able to retrieve sentence context to reject lures, whereas older adults can only use context to reject lures when global gist processing is disrupted during encoding by divergent sentences.

Although the methodologies and age ranges between these studies differ markedly, what appears consistent is that sentence processing can fundamentally alter the nature of false remembering depending on the list structure. Sentences converging on the meaning of critical lures can increase false memories for younger children and those that diverge from the meaning can reduce false memories for older children and adults (Dewhurst et al., 2007; Howe & Wilkinson, 2011; Thomas & Sommers, 2005). However, given the developmental differences in local and global gist processing, along with the fact that recollective retrieval processes differ across age groups (Hashtroudi et al., 1989; Henkel et al., 1998; Lindsay et al., 1991), it is difficult to pinpoint the exact mechanisms by which sentence processing influences false memories. To control for these issues, the current study investigates sentence processing within younger adults only.

The present study

The primary goal of the present study was to examine how varying the semantic context of sentences influences false memories in younger adults when DRM associates are placed in the context of sentences. For each experiment, we placed associates in the context of sentences blocked by theme such that the last word in each sentence was an associate of a non-presented critical lure. In Experiment 1, the sentence structure allowed for meaningful processing of list items for half the blocks, whereas the other half did not. The sentence context in Experiment 2 converged on the meaning of the critical lure in one condition and diverged from the meaning in another condition, and false-alarm rates were compared to a condition in which words were encoded in isolation. Experiment 3 used stimuli other than those typically used in the DRM paradigm that converged on two different meanings of a homographic lure (Hutchison & Balota, 2005) to compare recognition of items studied in sentences to words studied in isolation.

Although semantic processing has been shown to increase memory errors in the DRM paradigm (e.g., Goodwin et al., 2001; Toglia et al., 1999), providing contextual information that differentiates items at encoding or presenting encoding instructions that direct attention to differences among stimuli can reduce false memories (e.g., Arndt & Reder, 2003; Thomas & Sommers, 2005). Based on the mechanisms described by AMT and FTT, individual sentences that converge on the theme of a critical lure (local gist) are a necessary prerequisite for the creation of false memories. That is, semantic processing of the studied associates should increase lure or gist activation, making it more likely for the lure to be false recognized later. Importantly, however, it is also important to consider the global structure of the lists. In the current study, we hypothesize that list structure that allows for meaningful organization of the items within a list (global gist) should elevate false memories compared to lists that disrupt this processing. In addition to the influence of false remembering, meaningful list structure may make recognition of studied items easier. Finally, because sentence structure that converges on the meaning of the critical lure makes it easier to identify the themes (e.g., Howe & Wilkinson, 2011), it is possible that false alarms to critical lures may be greater for sentences than words. However, it has also been shown that sentence context can facilitate recollective retrieval processes in younger adults, resulting in fewer false alarms (Thomas & Sommers, 2005).

Experiment 1

In Experiment 1 we investigated the influence of semantic processing on false memories by placing DRM associates in the context of sentences where for half of the sentence blocks

the structure allowed for meaningful processing of sentences (e.g., "Stephanie lay in bed") whereas the other half did not (e.g., "Stephanie shaped in bed"). The only thing that differed between the two sentence types was the verb (i.e., lay vs. shaped) linking the subject (Stephanie) with the associate (bed). For clarity, we will refer to the former type as "meaningful" and the latter as "meaningless." Our primary hypothesis was that critical lure false alarms would be greater for meaningful blocks. This could occur for two reasons. Semantic and deep processing has been shown to increase both veridical and false memories (Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toglia et al., 1999). Thus, meaningful sentences may be more likely to result in local gist extraction. Alternatively, meaningful sentences may facilitate noticing the relations among other items within the list, making it more likely for global gist to be extracted. In either case, meaningful sentences should result in greater false alarms relative to the meaningless sentences in which this processing is disrupted.

Methods

All research reported herein was conducted using appropriate ethical guidelines and was approved by the Institutional Review Board at the University of Georgia. We report all data exclusions (if any) and all manipulations. No a priori power analysis was conducted. A sample size of at least 30 participants per condition was selected for each experiment based on previous false-memory studies using similar manipulations to the current study (e.g., Dewhurst et al., 2007; Thomas & Sommers, 2005) and from prior research in our laboratory (Marsh et al., 2003) showing robust effects with this sample size. All data are available on the Open Science Framework: https://osf.io/kn7v3/?view_only=bea9dcc3632148e8a634f6ee129d3977

Participants A total of 32 undergraduate students from the University of Georgia volunteered in exchange for partial credit toward a course research requirement. Each participant was individually tested in sessions that lasted approximately 15 min.

Materials The experimental materials consisted of 16 themed lists taken from the Roediger et al. (2001) norms. We selected eight semantic associates from each list to create sentences for each non-presented critical lure. Eight meaningful and eight meaningless sentences were created for each theme (256 total sentences) for counterbalancing purposes. Other than the verb used, the sentence structure was consistent for both meaningful and meaningless sentences. In the meaningful sentences, the verb used allowed for meaningful processing of the sentence (e.g., "John visited the hospital") whereas in the meaningless sentences it did not (e.g., "John dealt the hospital"). Note that in both sentence types the last word "hospital" is associated with the non-presented critical lure "doctor."

Design and procedure The study session was blocked within subjects, such that a block of eight meaningful sentences from six themed lists alternated with a block of eight meaningless sentences from six others. The sentences within each block were randomly presented, and for half of the participants a meaningful block was presented first while a meaningless block was presented first for the others. These alternating blocks persisted until 12 blocks (six meaningful and six meaningless) were presented. Items from the other four DRM sentence lists served as new items during the test phase. The counterbalancing scheme ensured each list was presented an equal number of times as meaningful and meaningless sentences as well as new lures across participants. So, for one-third of the participants, the "doctor" list was presented in meaningful blocks and in meaningless blocks for another third. Furthermore, for one-third of the participants, the "doctor" list was not studied and served as new items during the test phase. This process occurred for each of the 16 lists.

The test phase consisted of 48 old items and 48 new items. Of the old items, four were taken from each of the 12 presented lists. The new items consisted of 12 non-presented critical lures (e.g., "doctor") from each sentence list of the study phase. As described in the counterbalancing scheme, four sentence lists were not presented during study. Of these four lists, the critical theme word and four associates of each (20 items) were presented as new items. In addition, 16 new items were taken from the Roediger et al. (2001) norms that were unrelated to the other items. The 96 items during the test phase were randomly presented.

For each phase of the experiment, participants read the instructions from the computer monitor, which the experimenter also reiterated in her own words. The instructions for the study phase indicated that we were interested in seeing how people rated sentences for meaning. Participants were told that they would be presented with a series of sentences and to rate each sentence for subjective meaning on a scale from 1 to 7 (1 being absolutely meaningless, 7 being absolutely meaningful). The presentation rate was self-paced with a 5-s break between each block. Upon conclusion of the study phase, a 2-min distracter phase consisting of a series of mazes to be solved was administered. Following this, instructions for the surprise recognition test were given. Participants were told they were going to be shown a series of items. Upon presentation, they were to think back to the sentences rated earlier and if they remembered seeing the presented word in one of the sentences, they were to press the "yes" key. If the item was new, they were to press the key labeled "no" to indicate that they did not see the word during the previous rating task.

Results

Sentence rating task Table 1 displays means and standard deviations for meaningfulness ratings of stimuli in all three

experiments. Unless otherwise specified, all statistical tests are significant at the conventional 5% probability of a Type I error. To examine whether our encoding manipulation resulted in differences in perceived meaning between sentence types, a simple comparison was conducted for mean rating scores (1–7) for meaningful versus meaningless blocks. Meaningful blocks received significantly higher ratings than meaningless blocks, $F(1,31) = 326.99, p < .001, \eta_{p2} = .91$, suggesting the encoding task was successful in producing differences in perceived meaning.

Recognition The upper portion of Table 1 displays hits, false alarms to critical lures, and different categories of new items for Experiment 1. To account for overall propensity to respond “old,” we employed a correction for hits and false alarms to critical lures (see Kensinger & Schacter, 1999; Thomas & Sommers, 2005). For correct recognition, we subtracted the hit rate for meaningful and meaningless associates by the false-alarm rate for unstudied associates (which were related to different unstudied themes). For false recognition, we subtracted the false-alarm rate for meaningful and meaningless critical lures by the false-alarm rate to unstudied theme lures (which were related to the different unstudied associates). Note that this correction does not influence the critical comparison for the meaningfulness of items, as there are no “meaningful” or “meaningless” unstudied associates or themes. However, in subsequent experiments this correction is relevant so we thought it was most prudent to remain consistent across experiments.

Corrected hit rates to studied associates and false-alarm rates to critical lures were submitted to a 2 (block: meaningful

vs. meaningless) \times 2 (item type: studied vs. critical lure) repeated-measures analysis of variance (ANOVA).

There was no main effect of block, $F(1,31) = 3.29, p = .08, \eta_{p2} = .10$. A main effect of item type was found, $F(1,31) = 4.26, p = .048, \eta_{p2} = .12$, with greater recognition for studied items than critical lures. In addition, there was a significant interaction of block and item type, $F(1,31) = 5.26, p = .03, \eta_{p2} = .15$. False alarms to critical lures were greater in meaningful blocks than meaningless blocks, $F(1, 31) = 5.27, p = .03, \eta_{p2} = .15$. However, there were no significant differences in the hit rates for meaningful and meaningless blocks, $F(1, 31) = .04, p = .84, \eta_{p2} = .001$.

Discussion

The primary goal of Experiment 1 was to determine how varying semantic context influences false memories in the DRM paradigm and whether differences in false-alarm rates could be due to differences in gist processing. Processing of meaningful sentences increased false memories relative to meaningless sentences, whereas there were no significant differences between the two types of sentences in veridical memory. One possibility is that processing of meaningless sentences may have reduced the amount of semantic information extracted by the participants, thereby causing a reduction in false alarms. This idea is supported by differences in ratings scores between the two types of sentences. With decreased semantic processing in the meaningless blocks, the local gist representation of the critical lure may not have been as strongly activated as with the meaningful sentences. Alternatively, meaningful sentences resulted in relational processing that

Table 1 Recognition hit rates, false-alarm rates (standard errors), and meaningfulness ratings (standard deviations) for Experiments 1–3

Experiment and condition	Hits (associates)		False alarms (critical lure)		False alarms (completely new items)			Meaningfulness ratings	
	Studied	Corrected	CL	Corrected	Themed associates	Themes	Unrelated		
Experiment 1*									
Meaningful	.74 (.02)	.50 (.03)	.69 (.04)	.43 (.05)				5.87 (.69)	
Meaningless	.73 (.03)	.49 (.03)	.59 (.04)	.33 (.05)				2.14 (.75)	
Overall					.24 (.02)	.26 (.04)	.22 (.03)		
Experiment 2									
Convergent	.76 (.02)	.60 (.02)	.73 (.03)	.56 (.04)	.16 (.03)	.17 (.03)	.24 (.02)	4.33 (.85)	
Divergent	.69 (.02)	.39 (.04)	.49 (.04)	.04 (.05)	.30 (.03)	.45 (.04)	.33 (.03)	4.43 (.72)	
Word-only	.92 (.01)	.88 (.01)	.58 (.04)	.52 (.05)	.04 (.01)	.06 (.02)	.04 (.01)	4.35 (.46)	
Experiment 3									
Sentence	Grouped	.75 (.02)	.59 (.02)	.67 (.02)	.46 (.03)	.16 (.02)	.26 (.04)	.26 (.04)	4.09 (.66)
	Mixed	.73 (.02)	.58 (.03)	.61 (.03)	.43 (.03)	.15 (.02)	.28 (.04)	.28 (.04)	4.16 (.76)
Word	Grouped	.95 (.03)	.85 (.03)	.43 (.03)	.33 (.03)	.09 (.02)	.15 (.03)	.08 (.02)	4.18 (.67)
	Mixed	.96 (.01)	.87 (.02)	.44 (.04)	.35 (.03)	.10 (.02)	.19 (.04)	.07 (.01)	4.30 (.57)

Note. *Experiment 1 is within subjects, thus condition refers to study block
CL = critical lure

was disrupted in meaningless sentences, thus causing participants to focus more on item-specific processing during the latter. Item-specific processing has been shown to reduce false memories relative to relational processing, whereas no differences occur between the two types of processing for veridical memory (McCabe et al., 2004). It could be argued that veridical memory should be better when using item-specific processing. However, relational processing can serve as an effective means of discriminating between old and new items by responding "old" to items that are consistent with the gist of the studied items at the cost of increased false-alarm rates. Consistent with this idea, critical lures had a higher false-alarm rate when presented in the context of meaningful sentences presumably because they were consistent with the global gist of the other studied items. However, because meaningless processing may have disrupted both semantic extraction (local gist) and relational processing (global gist), we cannot arbitrate between these two alternatives. Experiment 2 was designed to examine whether relational processing will increase false alarms to critical lures while holding semantic processing constant (as indexed by meaningfulness ratings).

Experiment 2

Previous research has shown that sentences that diverge from the meaning of the critical lure eliminate age differences in false remembering (Howe & Wilkinson, 2011; Thomas & Sommers, 2005; for similar results with word pairs, see Odegard et al., 2008). It is suggested that divergent sentences disrupt relational processing during encoding, making it less likely for the global gist to be extracted. Therefore, we designed a similar experiment to that of Thomas and Sommers (2005) using a between-subjects design with convergent sentence (e.g., "After work he lay down on the bed"; "He had a frightening dream"), divergent sentence (e.g., "She walked along the river bed"; "The new car drives like a dream"), and word-only (e.g., bed, dream) conditions. Because sentences and words in this experiment are equally "meaningful" (as opposed to the meaningless sentences in the previous experiment), differences in false-alarm rates may be interpreted more precisely because local gist processing is held constant. Because convergent sentences elicit the meaning of the critical lures and are all related to one another within each list, we hypothesized that false alarms to critical lures would be greater than in the divergent condition in which sentences do not elicit the meaning of the critical lures and are unrelated to each other within each list. The inclusion of the word-only condition allows us to determine whether convergent sentences increase false memories, if divergent sentences decrease memories,

or both (Howe & Wilkinson, 2011). Thomas and Sommers (2005) found that for younger adults, false recognition was reduced in convergent compared to word-only conditions. However, this difference was considerably greater in divergent conditions. This suggests that convergence does not necessarily increase memories, but rather divergence decreases memories.

Methods

Participants Undergraduate students from the University of Georgia volunteered in exchange for partial credit toward a course research requirement. Each participant was individually tested in sessions that lasted approximately 20 min. Ninety new participants were randomly assigned to the convergent ($N = 30$), divergent ($N = 30$), or word-only ($N=30$) condition.

Materials A total of 12 themed lists with eight sentences in each were created with identical non-presented critical lures for each condition. The materials for the convergent and divergent sentences were borrowed from Thomas and Sommers (2005; we thank the authors for providing us with their stimuli). However, we slightly altered the sentences by trying to equate sentence length and eliminating proper nouns. Convergent sentences elicited the meaning of the semantic associates and converged on the meaning of the non-presented critical lure (e.g., "After work he lay down in bed"). Divergent sentences elicited a particular meaning of the associate at the end of each sentence, but did not converge on the meaning or gist of the non-presented critical lure (e.g., "She walked along the river bed"). Note that the last word of both types of sentences is an associate of the DRM theme word "sleep." In the word condition, the same associates were used as in the convergent and divergent conditions (e.g., "bed") but were presented in isolation (i.e., no sentences).

Design and procedure The procedure used in Experiment 2 was similar to Experiment 1, except that the 12 blocks were randomly presented during the study phase with each sentence (or word) within a block presented randomly. Instructions for the study and test phase in each condition were identical to those given in Experiment 1. After making meaningfulness ratings on all 12 blocks of sentences or words, participants engaged in a 2-min distractor phase and then were given instructions for the test phase. The test phase consisted of 48 old items and 48 new items randomly presented. In each condition, four old items were taken from each of the studied lists. The new items in all conditions consisted of the 12 non-presented critical themed items, as well as four associates from four themes that were never studied along with the critical lure from each. There were also 16 unrelated new items taken from other DRM lists.

Results

Sentence-rating task During the encoding task, there were no significant differences in ratings for meaning across conditions, $F < 1$, suggesting that the stimuli in one condition were not perceived as any more "meaningful" than another condition (see Table 1).

Recognition The middle portion of Table 1 displays hits, false alarms to critical lures, and false alarms to new unrelated items for Experiment 2. Due to differences in false responding to unrelated new items across conditions, $F(2, 87) = 51.52, p < .001, \eta_{p2} = .54$, we employed a correction for veridical recognition by subtracting the false-alarm rates to new associates from the hit rates (see Kensinger & Schacter, 1999; Thomas & Sommers, 2005). For false recognition, we subtracted the false-alarm rate to new themes from the false-alarm rate to critical lures (see Table 1). We conducted a 2 (item type: studied vs. critical lure) \times 3 (condition: convergent vs. divergent vs. word-only) mixed ANOVA. The analysis of corrected hit and false recognition scores revealed a main effect of item type, $F(1, 87) = 65.63, p < .001, \eta_{p2} = .43$, whereby critical lures were recognized less than studied items. A main effect of condition was also found, $F(2, 87) = 83.10, p < .001, \eta_{p2} = .66$. Participants in the divergent condition recognized fewer items than both the convergent and word-only conditions. These main effects were qualified by a significant interaction, $F(2, 87) = 12.60, p < .001, \eta_{p2} = .23$.

Separate ANOVAs were conducted for corrected hit and false-alarm rates across conditions. There was a significant difference in studied items recognized across conditions, $F(2, 87) = 86.05, p < .001, \eta_{p2} = .66$. Participants in the word-only condition recognized more studied items relative to the convergent condition, $F(1, 58) = 106.60, p < .001, \eta_{p2} = .65$, whereas fewer items were recognized in the divergent condition than in the convergent condition, $F(1, 58) = 21.52, p < .001, \eta_{p2} = .27$. There was also a significant difference in false alarms to critical lures between conditions, $F(2, 87) = 37.38, p < .001, \eta_{p2} = .46$. Participants in the divergent condition falsely recognized significantly fewer critical lures than the convergent condition, $F(1, 58) = 62.01, p < .001, \eta_{p2} = .52$. However, there were no significant differences between the word-only and convergent conditions, $F(1, 58) = .54, p = .47, \eta_{p2} = .01$.

Discussion

The purpose of Experiment 2 was to examine how sentence context influences global gist extraction compared to standard word processing. There was better veridical recognition in the word-only condition than in both sentence conditions, which should not be surprising due to the increasing demands of processing and storage (and subsequent remembering) of

sentences relative to words encoded in isolation (Thomas & Sommers, 2005). Participants in the convergent condition were also more likely to recognize studied items than participants in the divergent condition. Of critical interest was false recognition of critical lures, which was much lower in the divergent condition relative to the convergent and word-only conditions, which did not differ from each other. These results are consistent with the idea that participants are able to identify thematic associations when lists are organized in such a way to increase relational processing (Dewhurst et al., 2007), with this processing being disrupted when sentence themes diverge within lists. This relational processing may also have facilitated veridical recognition by responding "old" to items that are consistent with the global gist of the list. Because sentences in the divergent condition were dissimilar from one another within a list and not related to the critical lure, participants were unable to relate sentences together, and use shared cues to recognize studied items or falsely recognize critical lures.

Although we did not find evidence, as did Thomas and Sommers (2005), that younger adults were able to use sentence contexts to reduce false recognition in both sentence conditions compared to the word-only condition, there were several methodological differences that will be elaborated upon in the *General discussion*. However, these results are consistent with their findings in that the reason for the differences between the two sentence conditions is not because converging on the meaning of the critical lure increases false memories, but rather divergent sentences appear to drastically reduce false memories. To further examine the mechanisms of production of global gist, in Experiment 3 we disrupted relational processing by using sentences that converged on two different meanings of a homophone and presenting those sentences in either a grouped or mixed fashion.

Experiment 3

Prior research has shown that blocked presentation increases false alarms more so than random presentation (Brainerd et al., 2003; McDermott, 1996; Toglia et al., 1999). Furthermore, Goodwin et al. (2001) found that the grouping of unrelated filler items influenced false memories. For example, for the lure "soft" in one list, eight semantically related associates (e.g., hard, light, etc.) were presented followed by eight filler items unrelated to the critical (but related to the associate – hat, bulb, etc.). Other list structures alternated between related and unrelated items in groups of four, two, or one. It was found that false memories monotonically decreased as associative grouping decreased. Because local gist processing was equated across list format, this suggests the formatting disrupted global gist processing. Interestingly, a verbal "think aloud"

procedure revealed that participants sometimes extracted multiple gists.

In the current experiment, we used a similar approach using associative items other than those typically used in the DRM procedure. Stimuli were created by placing associates in the context of sentences that converged on two separate meanings of a non-presented critical homographic lure. For example, the word "fall" can refer to a season or to the act of stumbling. Associates were placed in the context of sentences that converged on each meaning (e.g., "The flowers bloomed in the spring" vs. "The slick ice caused her to slip"). In the *grouped* condition, four sentences from one meaning of the homographic lure (e.g., the season) were presented in succession and then sentences from the four sentences from the other meaning (e.g., to stumble) were presented (*grouped* condition). In the *mixed* condition, presentation of sentences was alternated between the two meanings. These were compared to similar word-only conditions in which words were presented in isolation. Based on findings from Goodwin et al. (2001), we anticipated that lure false alarms would be greater in the grouped than in the mixed condition, as mixed presentation may make it more difficult to extract the global gist of the list. However, because with this procedure there are technically two gists (e.g., season, stumble), it may be that relational processing is disrupted regardless of presentation format. Two word-only conditions were included to determine whether the increased semantic context created by sentences influenced the results.

Methods

Participants Undergraduate students from the University of Georgia volunteered in exchange for partial credit toward a course research requirement. Each participant was individually tested in sessions that lasted approximately 20 min. 150 new participants were randomly assigned to the sentence-grouped (N=41), sentence-mixed (N=35), word-grouped (N=38), or word-mixed (N=36) condition.

Materials A total of 12 homograph lists were used from Hutchison and Balota (2005) based off of the Twilley et al. (1994) norms. For the two separate meanings of the critical homograph, four associates were taken from each list. For example, for the homograph "fall", four words were related to the "autumn" meaning (e.g., "autumn," "season," "spring," "leaves") and four words were related to the "stumble" meaning (e.g., "stumble," "slip," "rise," "trip") to compose the eight-item list. The average backward associative strength (BAS) from each word to the critical homograph was equated between each meaning. We also created sentences for each of the 12 homograph lists, with four sentences related to one meaning (e.g., "The young boy hated raking leaves."), and

four sentences related to the other meaning (e.g., "The slick ice caused her to slip.") of the lure.

Design and procedure The only difference between the two sentence conditions was the order of presentation. In the *sentence-grouped* condition, the four sentences from one meaning were presented in succession, and then the four sentences from the alternate meaning were presented in succession. In the *sentence-mixed* condition, the sentences from the two different meanings were presented in alternating fashion. The same structure followed for the two word conditions; however, the associates were presented in isolation (i.e. no sentences). In the *word-grouped* condition, the four associates from each meaning were presented in succession, and in the *word-mixed* condition presentation of the associates alternated between meanings.

The procedure used in Experiment 3 was nearly identical to Experiment 2. However, in the grouped conditions, the four sentences (or words) from one meaning were randomly presented and then the four from the alternate meaning were presented, counterbalanced across participants as to which meaning was presented first. In the mixed conditions, the stimuli from the two different meanings were randomly selected to be presented in alternating form. After making meaningfulness ratings on all 12 blocks, participants engaged in a 2-min distractor phase and then were given instructions for the test phase. Instructions for the study and test phase in each condition were identical to those given in Experiments 1 and 2.

The test phase consisted of 48 old and 48 new items randomly presented. In each condition, two old items were taken from each meaning of the studied list (resulting in four items per list). The new items in all conditions consisted of the 12 non-presented critical homographs, as well as four homographs taken from the norms that were never studied, with three associates from each list. There were also 16 unrelated new items taken from other homograph lists.

Results

Sentence rating task Meaningfulness ratings during encoding were submitted to a 2 (context: sentence vs. word) \times 2 (presentation: grouped vs. mixed) between-subjects ANOVA. This revealed no effect of context, $F(1, 146) = 1.14, p = .29, \eta_{p2} = .01$, no effect of presentation, $F < 1$, and no interaction between the two, $F < 1$. This suggests that stimuli in one condition were not perceived as any more "meaningful" than another condition.

Recognition The lower portion of Table 1 displays hit rates, false-alarm rates to critical lures, and false-alarm rates to unrelated lures for Experiment 3. As with Experiment 2, we employed a correction for hits and false alarms to critical lures due to differences in false alarms to unrelated lures across

conditions, $F(3, 146) = 17.46$, $p < .001$, $\eta_{p2} = .26$. We conducted a 2 (item type: studied vs. critical lure) \times 2 (context: sentence vs. word) \times 2 (presentation: grouped vs. mixed) mixed ANOVA for average recognition. The analysis of corrected hit and false recognition scores revealed a main effect of item type, $F(1, 146) = 376.05$, $p < .001$, $\eta_{p2} = .72$, whereby critical lures were recognized less than studied items. There was no effect of presentation, $F(1, 146) = .017$, $p = .90$, $\eta_{p2} < .001$. A main effect of context was found, $F(1, 146) = 16.32$, $p < .001$, $\eta_{p2} = .10$, whereby more items were labeled "old" during word-only encoding than during sentence encoding. There was also a significant interaction of item type and context, $F(1, 146) = 125.20$, $p < .001$, $\eta_{p2} = .46$.

To examine the two-way interaction, separate ANOVAs were conducted for corrected hit and false-alarm rates between the two context conditions (collapsed across presentation formation). There was a significant difference in veridical recognition between the two encoding contexts, $F(1, 148) = 107.68$, $p < .001$, $\eta_{p2} = .42$, whereby participants in the word-only conditions recognized more studied items than the sentence conditions. There was also a significant difference in false alarms to critical lures between conditions, $F(1, 148) = 14.64$, $p < .001$, $\eta_{p2} = .09$. This comparison revealed that participants in the sentence conditions false alarmed to critical lures more often than participants in the word-only conditions. Thus, participants in the word-only conditions not only recognized more studied items than the sentence conditions, but also falsely recognized fewer critical lures.

Discussion

The purpose of Experiment 3 was to examine the influence of presentation format on false remembering when sentences or words converged on two separate meanings of a homographic lure. As with Experiment 2, we found that veridical recognition was better for word-only relative to sentence encoding conditions. In contrast, we found significant differences in false alarms to critical lures between sentence and word-only conditions that converged on the meaning of the critical lure, with greater false memories in the sentence conditions. Importantly, the manipulation to reduce relational processing by alternating presentation of homographic meaning failed to produce any differences in hits or false alarm within the sentence conditions or word-only conditions. One notable difference in the Goodwin et al. (2001) study is that their mixed format contained unrelated filler items that biased meaning away from the critical lure. In the current study, presentation alternated between two different meanings of the same homographic lure. It is possible that having to construct multiple gists (one for each meaning) may have disrupted relational processing regardless of presentation format. Thus, as in Experiment 2, diverging on the meaning of the critical lure may have reduced the production of shared cues that related

sentences (or words) together regardless of presentation format.

The finding that sentences led to greater false memories than words is inconsistent with Experiment 2. One possibility for the discrepancy is that in the current study, within each sentence there was often multiple pieces of information converge on the critical lure. For example, in the sentence, "The slick ice caused her to slip," the words, "slick," "ice," and "slip" could activate the critical lure "fall." Similarly, "The young boy hated raking leaves" has multiple pieces of information that activate "fall." In the word-only conditions, only "slip" and "leaves" would activate the critical lure. Robinson and Roediger (1997) found that increasing the number of associates within lists increased the probability of false recall, and a similar mechanism could be influencing our results. The rich semantic representations in the sentence conditions that converge on the critical lure may have therefore facilitated local gist extraction.

General discussion

Across three experiments we examined the influence of semantic context in the processing of DRM associates embedded in sentences. Previous research suggests that semantic processing influences false recognition by strengthening semantic relationships among items, making it more likely that the gist trace will be activated (Toglia et al., 1999). However, making stimuli more distinctive by providing contextual information or encoding instructions that direct attention to differences among stimuli serves to reduce false memories (e.g., Goodwin et al., 2001; McCabe et al., 2004). The present study demonstrates that additional contextual information does not necessarily reduce false memories and can actually increase false memories in the DRM paradigm, depending on the semantic properties of the stimuli. We provided contextual information that could be used to discriminate old from new items by placing DRM associates in sentences, finding that false memories were governed by the semantic properties of the stimuli that allowed for meaningful organization based on the similarities of the items.

Experiment 1 demonstrated that the meaning elicited by encoded stimuli influenced false memories. That is, the more meaningful items were perceived (as indicated by subjective ratings) the greater the false-alarm rates. Presumably, participants were able to form stronger relationships among items when processing allowed for more meaningful comprehension of the sentences. Experiments 2 and 3 extended these findings by suggesting that it is not simply how meaningfully the items are perceived that influences false memories (subjective ratings were equivalent across conditions), but rather the ability of the inherent properties of the stimuli to produce both relationships among studied items and connections from

the items to the critical lure. Thus, organization of items in the DRM paradigm that allow for meaningful relational processing of items within-lists and that converge on the semantic meaning of the critical lure increases the likelihood that the list theme is identified, resulting in more errors at test.

For both meaningful and convergent sentences/words, not only are the stimuli related to other items within-lists (global gist), but they also converge on the meaning of the critical lure (local gist). We believe that global gist improves veridical memory by increasing semantic relationships among items within a list while also increasing the probability that the theme is identified. Local gist does not necessarily facilitate veridical recognition, but increases false recognition because stimuli elicit the meaning of the critical lure making it more likely for a gist representation to be formed. Furthermore, as demonstrated in Experiments 1 and 3, the *degree* of external convergence (i.e., meaningfulness or increased backward associative strength; Robinson & Roediger, 1997) may be important during the retrieval process in order for the critical lure to cue the episodic representations of the studied items. In contrast, meaningless and divergent sentences presumably reduce both local and global gist processing, making false recognition less likely. Regardless of the exact mechanisms involved, the results from the present study suggest that subjective organization imposed by the participant during encoding is influenced by the semantic context in which DRM associates are imbedded.

It should be noted that our incidental learning paradigm is different from many past DRM studies that use intentional learning (e.g., Thomas & Sommers, 2005). The purpose of the incidental encoding by using meaningfulness ratings was two-fold. First, we wanted to ensure that we used a task that encouraged participants to process the *entire* sentence. If participants were able to process the stimuli freely, or even intentionally, it is possible that they would have caught onto our intention and only processed the final associate at the end of the list. If that were the case, that would reduce the efficacy of comparing sentences to words. Second, we wanted to ensure that our meaningfulness manipulation in Experiment 1 worked. Indeed, participants did rate meaning differently across the two stimulus types. Notably, however, semantic orienting tasks have been shown to elevate both veridical and false recall compared to non-semantic orienting tasks (Thapar & McDermott, 2001; Toglietta et al., 1999). Categorically blocked, relative to randomized, list presentation produces similar effects (Payne et al., 1996). One possible consequence of this is that conditions with stimuli that were arguably less semantically structured (i.e., meaningless stimuli, word stimuli, or mixed presentation) might have received greater “boosts” in semantic processing than the more semantically structured list items (i.e., meaningful stimuli, sentence stimuli, and grouped format). This could in part explain why there were no differences in hit rates across stimuli in

Experiment 1, false memory between words and convergent sentences in Experiment 2, or false memory between mixed and grouped stimuli. While we realize this does not explain the entirety of the results, it is nonetheless important to consider how the orienting task might interact with stimulus processing, particularly when using semantically structured lists. Future research comparing semantic versus non-semantic orienting tasks, or incidental versus intentional encoding, will better elucidate the mechanisms underlying false memory for sentence information.

The results from Experiment 2 are inconsistent with the findings from Thomas and Sommers' (2005) study. However, there were several methodological differences between our study and theirs. First, we made their stimuli more homogenous by removing proper nouns and trying to equate the sentence length, which may have reduced the distinctiveness of items. They also had participants intentionally remember stimuli that may have allowed participants to focus more on item-specific information. Furthermore, the authors used a within-subjects manipulation in order to minimize the possibility that participants employed a “distinctiveness heuristic” (Dodson & Schacter, 2002; Israel & Schacter, 1997; Schacter et al., 1999), whereas we used a between-subjects design. Although the distinctiveness heuristic could explain the reduction of critical lure false alarms in the divergent condition, it should also predict similar reductions in the convergent condition. It is unclear why participants would adopt differential decision criteria for rejecting lures in the two sentence conditions because both classes of stimuli are arguably more distinctive than the word-only condition. Rather, we propose that under incidental learning conditions, participants organized information based on the similarities between the studied items and that the list theme was more consistent with the studied items in the convergent and word-only conditions. Processing of distinctive information in the current study may have resulted in a disruption of relational encoding thus decreasing both veridical and false recognition.

Theoretical mechanisms

Two prominent theories provide mechanistic accounts of false memories in the DRM paradigm. AMT assumes false memories occur because studying a list of related associates implicitly activates the critical lure via automatic spreading activation (Anderson, 1983; Collins & Loftus, 1975; Gallo & Roediger III, 2002; Gallo, 2013; Roediger & McDermott, 1995), whereas FTT suggests that participants extract the overall theme, or gist, of the study lists (Brainerd & Reyna, 2001, 2004; Brainerd et al., 2001). The activated lure or gist trace representation produces a strong feeling of familiarity at test, and errors occur when recollective monitoring processes fail. As described previously, however, both theories make largely similar predictions in DRM studies because items that

are associatively related are also usually semantically related. In general, both theories can largely account for the findings in the current study. However, there are a few issues that arise in both instances, which we describe below.

Turning first to FTT, this theory argues that gist extraction occurs due to emergent semantic properties from the list structure. FTT therefore readily accounts for the finding that list structure fundamentally alters the likelihood of false remembering. In fact, FTT has been previously applied to not only the standard DRM paradigm, but also to text comprehension more broadly (see Reyna et al., 2016, for a review). In the context of the current study, we have distinguished between local and global gist. Local gist reflects meaning extracted from stimuli considered in isolation, whereas global gist reflects the extraction of meaning from the relation among list items (Lampinen et al., 2006; Neuschatz et al., 2002; Odegard et al., 2008). The finding that meaningful and convergent sentences increased false alarms is consistent with the idea that the list structure that allowed for meaningful relations to be noticed among list items made it more likely for the global gist to be extracted. In these conditions, the critical lure then becomes a good cue for the gist memory (e.g., sleep). FTT can also explain the finding that veridical recognition was worse in sentence compared to word-only stimuli. For words, the tested associate (e.g., bed) is an exact match to the studied item (e.g., bed), making it more likely for verbatim details to be retrieved. In sentence conditions, however, the tested associate (e.g., bed) is not as strong a cue for the verbatim details (e.g., “After work he lay down on the bed”). However, it may nevertheless cue the sentence-level (or local) gist, which can still be an effective means to recognize the item as “old.” The one finding that FTT has difficulty in accounting for is that grouped versus mixed encoding had no influence on performance in Experiment 3. Despite the fact that sentences (and words) converged on two different meanings of the same homographic lure (e.g., fall), alternating between meanings should have disrupted noticing similarities among list items making it more difficult to extract either of the gist meanings (e.g., season or stumble). It is possible that having two different meanings in the same list regardless of format disrupted gist processing, but we admit this this is a post hoc interpretation of the results.

The finding that alternating format did not influence false alarms could be accounted for by AMT. Because sentences (and words) converge on the meaning of the critical lure, it should not matter whether list presentation is grouped or mixed because each individual stimulus should spread activation to the critical lure (Goodwin et al., 2001). In our view, this sentence level convergence is similar to the idea of local gist processing. For similar

reasons, AMT can also account for the finding that more meaningful and convergent sentences increase false memories. However, this requires that an additional contextual constraint be applied to the likelihood that the critical lure is activated (Lampinen et al., 2006). That is, assuming there is a one-to-one association between related items, the associate “bed” should prime “sleep” regardless of whether it is presented in convergent or divergent sentence. The associative activation theory suggests that when a word or concept is encountered, this representation spreads activation to related theme nodes within the mental lexicon, which can include perceptual, conceptual, and spatial features (Howe, 2005; Howe et al., 2009; Howe & Wilkinson, 2011). These theme nodes can be activated by experience with the task (e.g., studying several related sentences) or pre-existing associations in memory. Importantly, this theory suggests that because words contain multiple meanings, there are actually many-to-many associations among these words in the mental lexicon (e.g., bed and sleep). Thus, the likelihood that a critical lure is activated depends on the context in which the associates are encountered. It should be noted, however, that this idea is conceptually similar to the global gist trace interpretation of the results posited by FTT. Regardless of the exact mechanisms, the findings from the current study suggest that any theory of DRM errors (semantic or associative in nature) must account for the context in which list items are embedded that allow for meaningful associations to be formed within a list.

Conclusion

In sum, the present study demonstrated the importance of contextual organization during encoding by showing that false memories are governed by semantic properties of the stimuli and the ability to activate the related theme. Although previous research suggests that distinctive processing can reduce the occurrence of false memories, this is not always the case. When associates were presented in the context of sentences, participants were more likely to falsely accept critical lures as old when the context allowed for meaningful relational processing of items. *Local gist* influences false recognition by increasing the likelihood that a global gist is formed. *Global gist* improves veridical recognition by enhancing relational processing that increases the semantic relationships among items and also increases the probability that the theme to be identified. However, this also increases the likelihood that a critical lure will cue the gist trace at retrieval, making false memory more likely. Future work exploring these ideas in other contexts (e.g., eyewitness suggestibility, social contagion, etc.) may help better understand mechanisms underlying the creation of false memories.

Knowledge gained from these studies can be used to develop means to reduce memory errors across a variety of important domains.

Data availability All data are available on the Open Science Framework: https://osf.io/kn7v3/?view_only=bea9dcc3632148e8a634f6ee129d3977

References

- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal Of Verbal Learning And Verbal Behavior*, 22(3), 261-295. [https://doi.org/10.1016/S0022-5371\(83\)90201-3](https://doi.org/10.1016/S0022-5371(83)90201-3)
- Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, 48(1), 1-15. [https://doi.org/10.1016/S0749-596X\(02\)00518-1](https://doi.org/10.1016/S0749-596X(02)00518-1)
- Brainerd, C., Chang, M., & Bialer, D. (2020). Emotional ambiguity and memory. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0001011>
- Brainerd, C., Payne, D., Wright, R., & Reyna, V. (2003). Phantom recall. *Journal of Memory and Language*, 48(3), 445-467. [https://doi.org/10.1016/S0749-596X\(02\)00501-6](https://doi.org/10.1016/S0749-596X(02)00501-6)
- Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience.
- Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 11(5), 164-169. <https://doi.org/10.1111/1467-8721.00192>
- Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review*, 24(4), 396-439. <https://doi.org/10.1016/j.dr.2004.08.005>
- Brainerd, C. J., Wright, R., Reyna, V., & Mojardin, A. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2), 307. <https://doi.org/10.1037/0278-7393.27.2.307>
- Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychonomic bulletin & review*, 15(6), 1035-1053. <https://doi.org/10.3758/PBR.15.6.1035>
- Clark, H. H., & Clark, E. V. (1977). Psychology and language. <https://doi.org/10.1017/S0305000900001562>
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407. <https://doi.org/10.1037/0033-295X.82.6.407>
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology*, 58(1), 17-22. <https://doi.org/10.1037/h0046671>
- Dewhurst, S. A., Pursglove, R. C., & Lewis, C. (2007). Story contexts increase susceptibility to the DRM illusion in 5-year-olds. *Developmental Science*, 10(3), 374-378. <https://doi.org/10.1111/j.1467-7687.2007.00592.x>
- Dodson, C. S., & Schacter, D. L. (2002). Aging and strategic retrieval processes: Reducing false memories with a distinctiveness heuristic. *Psychology and aging*, 17(3), 405. <https://doi.org/10.1037/0882-7974.17.3.405>
- Gallo, D. (2013). *Associative illusions of memory: False memory research in DRM and related tasks*. Psychology Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & cognition*, 38(7), 833-848. <https://doi.org/10.3758/MC.38.7.833>
- Gallo, D. A., & Roediger III, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, 47(3), 469-497. [https://doi.org/10.1016/S0749-596X\(02\)00013-X](https://doi.org/10.1016/S0749-596X(02)00013-X)
- Glucksberg, S., & Danks, J. (1975). *Experimental Psycholinguistics*. John Wiley and Sons.
- Goodwin, K. A., Meissner, C. A., & Ericsson, K. A. (2001). Toward a model of false recall: Experimental manipulation of encoding context and the collection of verbal reports. *Memory & cognition*, 29(6), 806-819. <https://doi.org/10.3758/BF03196410>
- Greene, C., Nash, R., & Murphy, G. (2020). Misremembering Brexit: Partisan bias and individual predictors of false memories for fake news stories among Brexit voters.
- Hancock, T. W., Hicks, J. L., Marsh, R. L., & Ritschel, L. (2003). Measuring the activation level of critical lures in the Deese-Roediger-McDermott paradigm. *American Journal of Psychology*, 116(1), 1-14. <http://www.press.uillinois.edu/about/permission.html>
- Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and aging*, 4(1), 106. <https://doi.org/10.1037/0882-7974.4.1.106>
- Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General*, 127(3), 251.
- Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1195. <https://doi.org/10.1037/0278-7393.25.5.1195>
- Hicks, J. L., & Marsh, R. L. (2001). False recognition occurs more frequently during source identification than during old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2), 375.
- Holliday, R. E., Brainerd, C. J., & Reyna, V. F. (2011). Developmental reversals in false memory: Now you see them, now you don't! *Developmental Psychology*, 47(2), 442. <https://doi.org/10.1037/a0021058>
- Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science*, 16(12), 927-931. <https://doi.org/10.1111/j.1467-9280.2005.01638.x>
- Howe, M. L., & Wilkinson, S. (2011). Using story contexts to bias children's true and false memories. *Journal of Experimental Child Psychology*, 108(1), 77-95. <https://doi.org/10.1016/j.jecp.2010.06.009>
- Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language*, 60(2), 229-251. <https://doi.org/10.1016/j.jml.2008.10.002>
- Hutchison, K. A., & Balota, D. A. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous and unambiguous critical lures. *Journal of Memory and Language*, 52(1), 1-28. <https://doi.org/10.1016/j.jml.2004.08.003>
- Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic bulletin & review*, 4(4), 577-581. <https://doi.org/10.3758/BF03214352>
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychol Bull*, 114(1), 3-28. <https://doi.org/10.1037/0033-2909.114.1.3>
- Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology*, 16(3-5), 399-415. <https://doi.org/10.1080/026432999380852>
- Kimball, D. R., Muntean, W. J., & Smith, T. A. (2010). Dynamics of thematic activation in recognition testing. *Psychonomic bulletin & review*, 17(3), 355-361. <https://doi.org/10.3758/PBR.17.3.355>
- Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29(2), 133-159. [https://doi.org/10.1016/0749-596X\(90\)90069-C](https://doi.org/10.1016/0749-596X(90)90069-C)

- Lampinen, J. M., Leding, J. K., Reed, K. B., & Odegard, T. N. (2006). Global gist extraction in children and adults. *Memory*, *14*(8), 952-964. <https://doi.org/10.1080/09658210601008957>
- Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, *52*(3), 297-318. [https://doi.org/10.1016/0022-0965\(91\)90065-Z](https://doi.org/10.1016/0022-0965(91)90065-Z)
- Loftus, E. F. (1971). Memory for intentions: The effect of presence of a cue and interpolated activity. *Psychonomic Science*, *23*(4), 315-316. <https://doi.org/10.3758/BF03336128>
- Marsh, R. L., Hicks, J. L., Cook, G. I., Hansen, J. S., & Pallos, A. L. (2003). Interference to ongoing activities covaries with the characteristics of an event-based intention. *J Exp Psychol Learn Mem Cogn*, *29*(5), 861-870. <https://doi.org/10.1037/0278-7393.29.5.861>
- Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 886.
- McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic bulletin & review*, *11*(6), 1074-1079.
- McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language*, *35*(2), 212-230.
- Meade, M. L., Watson, J. M., Balota, D. A., & Roediger III, H. L. (2007). The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. *Journal of Memory and Language*, *56*(3), 305-320.
- Neuschatz, J. S., Lampinen, J. M., Preston, E. L., Hawkins, E. R., & Toglia, M. P. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *16*(6), 687-708.
- Odegard, T. N., Holliday, R. E., Brainerd, C. J., & Reyna, V. F. (2008). Attention to global gist processing eliminates age effects in false memories. *Journal of Experimental Child Psychology*, *99*(2), 96-113.
- Otgaar, H., Howe, M. L., Peters, M., Smeets, T., & Moritz, S. (2014). The production of spontaneous false memories across childhood. *Journal of Experimental Child Psychology*, *121*, 28-41.
- Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*(2), 261-285.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1-75.
- 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), 1-9. <https://doi.org/10.1016/j.jarmac.2015.12.003>
- Rhodes, M. G., & Anastasi, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic bulletin & review*, *7*(1), 158-162.
- Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*(3), 231-237.
- 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*(4), 803.
- 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*(3), 385-407.
- 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), 1-24.
- Swannell, E. R., & Dewhurst, S. A. (2013). Effects of presentation format and list length on children's false memories. *Journal of cognition and development*, *14*(2), 332-342.
- 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*(3), 424-432.
- Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language*, *52*(1), 71-86.
- Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, *7*(2), 233-256.
- Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & cognition*, *22*(1), 111-126.
- Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of experimental psychology*, *70*(1), 122.
- Unsworth, N., & Brewer, G. A. (2010). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language*, *62*(1), 19-34.
- Winograd, E. (1968). List differentiation, recall, and category similarity. *Journal of experimental psychology*, *78*(3p1), 510.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.