

List blocking and longer retention intervals reveal an influence of gist processing for lexically ambiguous critical lures

Mark J. Huff¹ · Jaimie McNabb² · Keith A. Hutchison²

Published online: 24 June 2015
© Psychonomic Society, Inc. 2015

Abstract In two experiments, we examined veridical and false memory for lists of associates from two meanings (e.g., *stumble*, *trip*, *harvest*, *pumpkin*, etc.) that converged upon a single, lexically ambiguous critical lure (e.g., *fall*), in order to compare the activation-monitoring and fuzzy-trace false memory accounts. In Experiment 1, we presented study lists that were blocked or alternated by meaning (within subjects), followed by a free recall test completed immediately or after a 2.5-min delay. Correct recall was greater for blocked than for alternated lists. Critical-lure false recall was greater for blocked lists on an immediate test, whereas both list types produced equivalent false recall on a delayed test. In Experiment 2, lists blocked and alternated by meaning were presented via a between-subjects design, in order to eliminate possible list-type carryover effects. Correct recall replicated the result from Experiment 1; however, blocking lists increased false recall on delayed, but not on immediate, tests. Across the experiments, clustering correct recall by meaning increased across the delay selectively for the alternated lists. Our results suggest that thematic (i.e., gist) processes are influential for false recall, especially following a delay, a pattern consistent with fuzzy-trace theory.

Keywords False memory · Lexical ambiguity · Blocking · Retention interval · ARC clustering

✉ Mark J. Huff
mhuff@wustl.edu

¹ Department of Psychology, Washington University, One Brookings Drive, CB 1125, St. Louis, MO 63130, USA

² Montana State University, Bozeman, MT, USA

The notion that learned mental representations are organized on the basis of their relations to other, similar representations enjoys a rich history within cognitive psychology (Collins & Loftus, 1975; Masson, 1995; Neely, 1991; see, too, Burgess & Lund, 2000, for a discussion). Information that shares a high number of associative links, occurs in similar temporal sequences, or is organized hierarchically is postulated to be grouped together within one's mental network (Anderson, 1983; Anderson & Bower, 1972; Kintsch, 1974). Effects of similarity have been shown to exert a strong influence on episodic memory. For instance, free recall is greater from semantically related word lists than from unrelated word lists (Huff, Meade, & Hutchison, 2011; Hunt & Einstein, 1981; Rabinowitz, Craik, & Ackerman, 1982), and participants are likely to cluster conceptually similar items together (Bousfield, 1953; Mandler, 1967; Zaromb & Roediger, 2010), even when the items are not organized by similarity at study (Cofer, 1975). Furthermore, clustering by similarity at retrieval is also found when participants study words using relational (vs. item-specific) processing tasks that facilitate an organizational structure at encoding (Hunt & Einstein, 1981; although see Huff & Bodner, 2014, for exceptions). Thus, increased processing of semantic relations through either encoding instructions or blocking the study materials by meaning increases participants' clustering of related items at recall.

Although increased processing of similarity is generally beneficial to memory, semantic relatedness can produce occasional memory errors (Underwood, 1965), such as in the powerful Deese/Roediger–McDermott (DRM) false memory paradigm (Deese, 1959; Roediger & McDermott, 1995). Using a homograph variant of this paradigm, our study evaluated the effects of thematic similarity in correct and false recall when study lists were blocked or alternated by meaning at study.

The DRM paradigm presents participants with study lists of associated words (e.g., *bedroom*, *wake*, *pillow*, etc.) that converge upon a single, nonpresented critical lure (CL; e.g., *sleep*) that is often falsely recalled or recognized on a later test. The DRM illusion is robust: False recall often approaches 50% (Roediger & McDermott, 1995), and false recognition has been shown to equal the hit rates for studied items (Lampinen, Neuschatz, & Payne, 1999). The illusion is also difficult to eliminate, persisting after explicit warnings (Gallo, Roberts, & Seamon, 1997; McCabe & Smith, 2002; McDermott & Roediger, 1998) and after study tasks that elicit distinctive processing (Huff, Bodner, & Fawcett, 2015; Hunt, Smith, & Dunlap, 2011; Israel & Schacter, 1997).

Two popular accounts of the DRM false memory illusion are the activation-monitoring theory (AMT; Roediger, Balota, & Watson, 2001) and fuzzy-trace theory (FTT; Brainerd & Reyna, 2002; Reyna, 1998; but see Miller, Guerin, & Wolford, 2011, and Miller & Wolford, 1999, for a decision-criterion account). AMT postulates that DRM false memories are created during study and retrieval through automatic spreading activation (Collins & Loftus, 1975) from the list item representations to the CL representation. This activation presumably summates and converges on the CL (Roediger et al., 2001a, b). The amount of associative activation summing on a given CL is operationalized in terms of its mean backward associative strength (BAS) from the list items to the CL, with BAS defined in published word association norms as the probability that a given list item will produce the CL (e.g., Nelson, McEvoy, & Schreiber, 1999). CL activation from list items is also suggested to occur at retrieval through a retrieval-mode process (Tulving, 1983), in which attempting to retrieve studied items reactivates the associative network of items (both studied and associated) established at study (Meade, Watson, & Balota, 2007). At retrieval, participants must engage in source monitoring to determine whether the activated items were actually studied, utilizing retrieved contextual details regarding the study episode (Johnson, Hashtroudi, & Lindsay, 1993). Therefore, according to AMT, the DRM illusion is due to associative activation, plus a monitoring failure in which participants fail to reject the CL.

In contrast, FTT posits that encoding creates two memory representations formed in parallel: verbatim and gist (Brainerd & Reyna, 1990). The *verbatim* representation contains perceptual and contextual details of an encoded item, whereas the *gist* representation contains the overall meaning of the item or grouping of items. At retrieval, participants can rely on stored verbatim or gist representations to report information from memory. When studying DRM lists, thematic relations between the studied items are encoded as a gist representation, which is similar in meaning to the CL itself. This similarity between the CL's meaning and the stored gist representation produces the false memory illusion at test (Brainerd, Payne, Wright, & Reyna, 2003; Brainerd & Reyna, 2002). Thus, the

DRM false memory illusion is attributed to a persistent gist representation and a concomitant reduction in the verbatim representation, due to interference or decay.

Separating the AMT and FTT accounts in the DRM paradigm has proven quite difficult, due to a confound in which DRM list items are both associatively related and similar in meaning to the CL. Deese (1959) reported that a list's mean BAS was highly correlated with CL false recall ($r = +.87$). Similarly, Roediger, Watson, McDermott, and Gallo (2001) reported that BAS was the best predictor of CL false recall and false recognition, accounting for 68% of the variance in recall and 48% of the variance in recognition. McEvoy, Nelson, and Komatsu (1999) also found that lists with higher BAS produced greater false recall of CLs. Thus, BAS from list items to CLs is highly predictive of the DRM illusion, consistent with the AMT predictions.

Separately, Brainerd, Yang, Reyna, Howe, and Mills (2008) found that, when using a principal components analysis, CLs rated in Toglia and Battig's (1978) semantic word norms as being highly familiar and meaningful loaded on the same factor as false recall and false recognition. The authors interpreted these loadings as evidence for FTT, suggesting that the DRM illusion was driven by a CL's semantic meaning, rather than by BAS from the list items to CLs. However, BAS also loaded on this same factor and, because a principal components analysis does not evaluate unique variance accounted for by individual variables, it is unknown whether BAS or CL familiarity and meaningfulness were stronger predictors of the DRM illusion.

To achieve greater control over the variables thought to influence false memory, researchers have compared AMT and FTT by manipulating the study materials to be high or low in either BAS or thematic gist. A simple way to do this is to present multiple lists in blocked or random order. Blocking lists by meaning should assist in constructing an overall thematic structure for each list that can guide later recall. Consistent with FTT, researchers have reported an increase in false memory when related list items were presented in a blocked versus a random order (Mather, Henkel, & Johnson, 1997; McDermott, 1996; Toglia, Neuschatz, & Goodwin, 1999).

Alternatively, rather than disrupting gist-based processing through random presentation, researchers have also designed study lists that lack thematic consistency. For instance, Huff and Hutchison (2011) presented participants with lists of unrelated words (e.g., *slope*, *reindeer*, and *corn*) that were related to nonpresented mediators (e.g., *ski*, *sleigh*, *flake*) that converged upon a nonpresented CL (e.g., *snow*). These lists were referred to as *mediated lists*, since they were indirectly associated with CLs through nonpresented mediators. The lack of a consistent meaning on mediated lists prevents the creation of a theme-based gist representation. Consistent with AMT, Huff and Hutchison (see, too, Huff, Coane, Hutchison, Grasser, & Blais, 2012) found that initial attempts to recall the list items

or guess the mediated CLs inflated false alarms of the mediated CLs on a final recognition test.

Similar to mediated false memory experiments, Cann, McRae, and Katz (2011) created lists in which study lists were organized by situational (i.e., script) knowledge (e.g., *band, fans, lighters*, etc., for *concert*). These lists were argued to only evoke a gist representation, because BAS was minimized (mean BAS = .06). The results showed that CL false alarms from the gist lists ($M = .65$) were similar to those from DRM lists ($M = .75$), suggesting that gist extraction can produce a comparable false memory effect. However, false recognition from these strong-gist lists ($M = .65$) was even more similar to that from Roediger, Watson, et al.'s (2001) standard DRM lists ($M = .60$), which were comparable in BAS (range = .02–.11) to Cann et al.'s gist lists. Thus, despite creating lists with strong gists, which should have increased false memory according to FTT, false recognition was similar to that from standard DRM lists with equally low BAS.

Similarly, associative activation and gist extraction have also been compared by varying the number of gist representations presented in a related list. Hutchison and Balota (2005) used two types of 12-item word lists that were equated in BAS to a CL, but that differed in the numbers of meanings presented. The first list type was a standard DRM list that converged upon a single meaning. The second list type utilized a homograph CL (e.g., *fall*), in which six list items converged upon one meaning of the CL (e.g., “stumble”), and six items converged upon another meaning (e.g., “autumn”; see Balota & Paul, 1996, for a similar procedure in a semantic-priming paradigm). It was hypothesized that if CL false memories were due to gist extraction, then increasing the number of study items from six to 12 items would only increase false memories for standard DRM, but not homograph lists. This is because increasing study items in homograph lists would add a second meaning that conflicted with the gist representation from the first meaning. In contrast, if DRM false memories were due to associative activation, false memories would increase similarly for DRM and homograph lists when six additional items were included, due to an equivalent increase in BAS. Across five experiments, Hutchison and Balota found that increasing the number of study items indeed increased false recall and recognition equally for both DRM and homograph lists, a pattern consistent with AMT. Furthermore, when participants were instructed to judge the similarity of meanings between CLs and their respective word lists, adding six associates increased the rated similarity more for DRM than for homograph lists. Thus, despite a greater increase in gist for DRM lists than for homograph lists, false memory rates remained equivalent across list types.

In a cross-experiment analysis, Hutchison and Balota (2005) also compared the effects of blocking (Exp. 1) versus alternating (Exp. 2) meanings in their homograph lists during presentation (see Table 1). Alternating meanings decreased veridical memory but had no effect on false memory, a pattern suggesting that maximizing gist-based processing through blocking does not inflate false memory beyond the level from alternating meanings. Notice, however, that this finding differs from the patterns described earlier, in which blocking generally increased false memory (Mather et al., 1997; McDermott, 1996; Toggia et al., 1999). One possibility for this discrepancy is that the past studies confounded blocking of meaning with blocking of BAS, whereas in Hutchison and Balota's lists, CL activation could still summate from sequentially adjacent items, even in the alternated-list condition. Previous blocking studies had also included longer retention intervals and a greater number of items and themes within each study list than did Hutchison and Balota. Longer retention intervals and increased numbers of items/themes could increase reliance on gist at recall to organize the studied information.

In reference to the retention interval difference between studies, a criticism of Hutchison and Balota's (2005) findings is that their results might generalize only to immediate recall. The DRM illusion can persist for weeks and even months after study (Seamon et al., 2002; Thapar & McDermott, 2001; Toggia et al., 1999), which may reflect contributions of different processes following a delay. In the DRM paradigm, increasing the retention interval decreases memory for studied items, but often does not decrease false memory for CLs (Payne, Elie, Blackwell, & Neuschatz, 1996; Thapar & McDermott, 2001). According to FTT, this pattern can be accounted for by a decay of verbatim but not gist memory, causing participants to rely more on a gist representation to retrieve information from memory over time. By blocking or alternating homograph lists by meaning in order to vary gist consistency, gist-based contributions can be evaluated both immediately and after a delay while holding both list items and BAS constant.

Aside from using only immediate tests, additional features of Hutchison and Balota's (2005) design were suboptimal for evaluating the effects of blocking. First, Hutchison and Balota examined blocking effects across separate experiments. A more ideal comparison would involve samples taken from the same participant population in the same experiment. Second, a reexamination of their lists revealed that the rank orderings of BAS differed across the blocked and alternated lists. In the blocked condition (Exp. 1), BAS had a scalloping pattern: descending BAS for the first six items, followed by descending BAS for the second six items. Alternating the sublists in Experiment 2, however, created a more consistent descending BAS pattern, in which

stronger items were presented early (see the top of Table 1). Given that items in the earlier portion of the list might produce greater implicit priming of CLs and

greater false memory (Meade, Hutchison, & Rand, 2010; Hutchison, Meade, Williams, McNabb, & Manley, 2015), elevated processing from the initially strong

Table 1 Sample homograph list orders, along with the backward associative strength for each list item, for the critical lure *fall*, used both by Hutchison and Balota (2005) and in the present experiments

	6- Item List	12- Item List (Blocked)	12- Item List (Alternated)
	Hutchison & Balota (2005)		
<i>"fall"</i>	STUMBLE(.70) SLIP (.51) RISE (.33) TRIP (.29) FAINT (.21) CLUMSY (.09)	STUMBLE (.70) SLIP (.51) RISE (.33) TRIP (.29) FAINT (.21) CLUMSY (.09) AUTUMN (.52) SEASON (.19) SPRING (.19) LEAVES (.17) BRISK (.00) HARVEST (.00)	STUMBLE(.70) AUTUMN (.52) SLIP (.51) SEASON (.19) RISE (.33) SPRING (.19) TRIP (.29) LEAVES (.17) FAINT (.21) BRISK (.00) CLUMSY (.09) HARVEST (.00)
	Or		
	Present Study		
<i>"fall"</i>	STUMBLE(.70) SLIP (.51) RISE (.33) TRIP (.29) FAINT (.21) CLUMSY (.09)	STUMBLE (.70) SLIP (.51) RISE (.33) TRIP (.29) FAINT (.21) CLUMSY (.09) AUTUMN (.52) SEASON (.19) SPRING (.19) LEAVES (.17) BRISK (.00) HARVEST (.00)	STUMBLE (.70) SEASON (.19) RISE (.33) LEAVES (.17) FAINT (.21) HARVEST (.00) SLIP (.51) AUTUMN (.52) TRIP (.29) SPRING (.19) CLUMSY (.09) BRISK (.00)
	Or		

items in alternated lists could mask any advantage for lists blocked by meaning.¹ Finally, for some lists, the Nelson et al. (1999) norms did not provide six associates for each meaning, so additional associates were normed but placed at the end of those lists. The BAS of these newly normed items was greater ($M = .135$) than that of the other items in the list ($M = .103$), and adding these items to the end of the list disrupted the BAS order. For example, the list item *scarlet* ($M = .48$) was placed at the end of the *letter* list, even though it had a stronger BAS than the first list item, *number* ($M = .17$). Therefore, to clearly examine gist and associative effects in false memory for homograph CLs, we sought to rectify the presentation differences between blocked and alternated lists.

The present experiments

In the present study, we sought to extend Hutchison and Balota's (2005) design to more strongly test the extent to which gist-based and association-based processing contribute to false memory. This extension included four improvements over Hutchison and Balota's methodology. First, to control for item differences across lists, the present study tested only homograph lists. Second, the alternated homograph lists were reorganized to reflect a BAS order similar the one in the blocked lists, in which the beginning of the second 6-word list included a high-BAS item (see the bottom of Table 1). Both list types now included in the BAS ordering the list items normed by Hutchison and Balota (see the asterisk items in the Appendices). Third, participants recalled lists either immediately following study or following a 2.5-min delay. Finally, blocking lists by meaning was manipulated both within subjects (Exp. 1) and between subjects (Exp. 2), to replicate Hutchison and Balota and to control for potential list type carryover effects.

We expected that CL false recall would be equivalent across the blocked and alternated lists on an immediate test, consistent with AMT and replicating Hutchison and Balota (2005). We also expected to find a significant blocking effect on immediate correct recall, showing that thematic-based processing is indeed important to veridical memory, but not to false memory. Following a delay, if false memories are still governed exclusively by associative activation, false recall of the CLs should remain equivalent for blocked and alternated lists. In contrast, if gist processes are stronger after a delay, as

the verbatim trace fades, then false recall should be selectively higher for the blocked lists than for the alternated lists. Finally, adjusted-ratio-of-clustering (ARC; Roenker, Thompson, & Brown, 1971) scores were calculated, so as to provide potential converging evidence of the presence of gist processing. The clustering of list items initially alternated during presentation should act as a marker of gist-based reliance at recall.

Experiment 1: Within-subjects manipulation of list type

Method

Participants A total of 145 Montana State University Psychology undergraduates participated for partial fulfillment of an introductory psychology research requirement.² The participants were randomly assigned to either the no-delay or delay conditions. All were native English speakers with normal or corrected-to-normal vision. No other demographic information was collected.

Design This experiment used a 2 (List Type: blocked vs. alternated) \times 2 (Delay: no delay vs. 2.5-min delay) mixed design in which List Type was manipulated within subjects and Delay between subjects.

Materials Fourteen of Hutchison and Balota's (2005) homograph word lists were used in the present study (see the Appendices). Each list contained six words related to one meaning (e.g., *stumble*, *slip*, etc.) and six words related to a second meaning (e.g., *autumn*, *season*, etc.), of which both meanings converged upon the same CL (e.g., *fall*). The lists were blocked or alternated by meaning (see the bottom of Table 1). The blocked lists contained six words related to the first meaning, followed by six words related to the second meaning, just as in Hutchison and Balota's study. However, unlike in Hutchison and Balota, their newly normed words were now inserted within the descending BAS order within each sublist. For the alternated lists, BAS was reorganized from Hutchison and Balota's alternated lists to more closely resemble the BAS order in the blocked lists. Specifically, because the list items in the blocked condition were presented in descending order for Sublist 1 (S1) followed by Sublist 2 (S2), we used the following formula for the alternated lists, in which the number following the sublist indicates the BAS order (1–6) of the list items within each sublist, S1-1, S2-2, S1-3, S2-4, S1-5, S2-6, S1-2, S2-1, S1-4, S2-3, S1-6, S2-5 (see the bottom of Table 1).

¹ However, this "front-loaded BAS" explanation is unlikely to be an explanation for Hutchison and Balota's (2005) null blocking effect for homograph lists, because they used the same blocked-versus-alternated manipulation for their DRM lists across Experiments 1 and 2 and also found no blocking effect. If higher initial BAS for alternated lists increased false memory, this should have occurred for the DRM lists.

² Experiment 1 originally consisted of two experiments, 1a ($N = 72$) and 1b ($N = 72$), with 1b being a direct replication of 1a. However, because the recall patterns were similar in both experiments, previous reviewers suggested that we combine them for a more efficient presentation of our results.

This created a similar scalloped order of BAS for both list types. Seven alternated lists (Appendix 1) and seven blocked lists (Appendix 2) were created for each participant. List order presentation was randomized, and blocking of the word lists was counterbalanced across subjects.

Procedure Participants were seated at a computer monitor and asked to pay attention and remember the words presented on the screen. The words were presented one at a time for 1, 200 ms (500-ms interstimulus interval) each, in lowercase letters in the center of the screen. Following study, the participants in both the delay and no-delay conditions were given 1 min to freely recall as many list words as possible by writing them down on a sheet of paper. In the delay condition, the study and recall phases were separated by an arithmetic filler task for 2.5 min, whereas those in the no-delay condition completed a recall test immediately following study. Participants repeated this procedure until all of the word lists had been shown and recalled. After all study–test trials, participants were debriefed and given credit.

Results

For all results reported, statistical significance was set at $p < .05$ unless otherwise noted. Effect sizes were calculated using partial eta-squared (η_p^2) for analyses of variance (ANOVAs) and Cohen's d for t tests for all significant findings. Correct recall, ARC clustering, and CL false recall were each analyzed using a 2 (Blocking: blocked vs. alternated list) \times 2 (Delay: no delay vs. delay) mixed ANOVA. The proportions of correct recall and CL false recall, along with ARC clustering scores, are reported in Table 2.

Table 2 List item recall, critical item recall, and adjusted-ratio-of-clustering (ARC) scores for Experiments 1 and 2

	Experiment 1		Experiment 2	
	No Delay	Delay	No Delay	Delay
List Items				
Blocked lists	.60 (.01)	.50 (.01)	.62 (.02)	.52 (.02)
Alternated lists	.54 (.01)	.46 (.01)	.59 (.02)	.51 (.02)
ARC Scores				
Blocked lists	.67 (.02)	.58 (.02)	.62 (.04)	.59 (.04)
Alternated lists	.05 (.03)	.23 (.03)	.13 (.06)	.30 (.06)
Critical Items				
Blocked lists	.34 (.02)	.40 (.02)	.17 (.03)	.33 (.05)
Alternated lists	.25 (.02)	.39 (.02)	.25 (.03)	.20 (.04)

In Experiment 1, we manipulated blocked and alternated list types within subjects, whereas in Experiment 2 we manipulated list types between subjects. Retention interval was manipulated between subjects in both experiments. Standard errors are listed in parentheses.

Correct recall Significant effects of blocking, $F(1, 143) = 46.82$, $MSE = .004$, $\eta_p^2 = .25$, and delay, $F(1, 143) = 30.64$, $MSE = .004$, $\eta_p^2 = .18$, were found. Correct recall was greater on lists blocked rather than alternated by meaning (.55 vs. .50), and greater when testing was immediate rather than delayed (.57 vs. .48), consistent with previous work (Hutchison & Balota, 2005; Payne et al., 1996). The Blocking \times Delay interaction approached, but did not reach, significance, $F(1, 143) = 2.78$, $MSE = .004$, $p = .10$, $\eta_p^2 = .02$.

ARC analysis One possible method for examining the contributions of gist-based processing at test is by calculating ARC indices for correctly recalled list items (Roemer et al., 1971; Senkova & Otani, 2012). These scores assess the extent to which items reported during free recall are organized by category membership, which, for our purposes, acts as a metric for determining thematic organization at retrieval (1 = perfect clustering by meaning, 0 = chance clustering, -1 = perfect alternating of meaning). In this case, a greater reliance on gist processing would be shown by greater positive ARC scores. An ARC score was calculated for each list recalled by participants. An inclusion requirement was that participants had to recall at least one item from each of the two meanings presented in the study list. If participants failed to recall at least one item from each of the two meanings, those lists were not included in calculating the mean ARC score for each list type. Given this criterion, ARC scores were computed for 96% of the lists in the no-delay group, and 86% of the lists in the delay group. Thus, ARC scores were available for analysis from a large majority of the lists. Extralist intrusions and CLs that were falsely recalled were not entered into the ARC calculation for a given list.

For the analyses of ARC clustering, a main effect of blocking was found, $F(1, 143) = 187.39$, $MSE = .09$, $\eta_p^2 = .57$, revealing that clustering was greater for lists blocked than for those alternated by meaning (.62 vs. .14). The effect of delay was not significant, $F(1, 143) = 2.16$, $MSE = .08$, $p = .14$, but the Blocking \times Delay interaction was significant, $F(1, 143) = 14.28$, $MSE = .09$, $\eta_p^2 = .09$. Follow-up tests revealed that, for blocked lists, clustering was marginally greater on immediate than on delayed tests (.67 vs. .58), $t(143) = 1.95$, $SEM = .04$, $p = .05$, $d = 0.34$, but for alternated lists, clustering increased between the immediate and delayed tests (.05 vs. .23), $t(143) = 3.50$, $SEM = .04$, $d = 0.58$, suggesting that the use of gist information was greater for alternated lists following a delay.

False recall Significant effects of blocking, $F(1, 143) = 5.48$, $MSE = .03$, $\eta_p^2 = .04$, and delay, $F(1, 143) = 11.97$, $MSE = .07$, $\eta_p^2 = .08$, were also found on CL false recall. False recall was greater for blocked than for alternated lists (.37 vs. .32), and greater on a delayed than on immediate test (.40 vs. .29). The interaction was also significant, $F(1, 143) = 4.20$, $MSE = .03$, $\eta_p^2 = .03$. Unexpectedly, however, on an immediate test, false recall was greater on blocked than on alternated lists (.34 vs.

.25), $t(71) = 3.36$, $SEM = .03$, $d = 0.47$, however, on a delayed test, false recall on alternated lists increased to match that on blocked lists (.39 vs. .40).

The total numbers of extralist intrusions per list were also analyzed, though they were rarely falsely recalled ($M = .20$), and the majority of these intrusions were related to the meanings presented in the list ($M = .17$). We observed no differences in the numbers of extralist intrusions per list between list types or delays, and no interaction was found, $F_s < 1$.

Discussion

For correct recall, blocking lists by meaning at study increased recall, presumably due to a strengthened semantic list structure at encoding. Furthermore, correct recall decreased across a delay, consistent with general forgetting patterns (Payne et al., 1996). Both of these findings were expected. Consistent with these patterns, ARC clustering was much greater for blocked than for alternated lists, particularly on an immediate test. Clustering also interacted with retention interval, such that a trend was found for clustering to be lower for blocked lists, but greater for alternated lists, following a delay. The latter pattern suggests that participants may be more reliant on gist information after a delay when thematic information is disrupted on alternated lists. Of course, this alone is not necessarily evidence for greater semantic processing, because recalling based on serial presentation order would also produce large clustering differences between list types. Importantly, however, the ARC analysis suggests that participants became less reliant on the verbatim serial-order presentation on both blocked and alternated lists after a delay, a finding consistent with FTT.

For false recall, however, blocking during study produced an increase in false recall over alternated lists on an immediate test, but both list types were equivalent in false recall on a delayed test. The greater increase in false memory over delay for alternated lists is consistent with the ARC clustering scores in suggesting that participants used thematic gists to reorganize the items from the alternated lists following the delay. This increase in gist-based retrieval might, in turn, have caused an increase in false recall.

The immediate test results are inconsistent with those of Hutchison and Balota (2005), who found that blocked and alternated lists in cross-experimental comparisons produced equivalent false recall on immediate tests. One explanation for this difference may be our use of a within-subjects blocking design, which is in contrast to the between-subjects design used by Hutchison and Balota. Using a within-subjects manipulation, exposure to both list types might have led participants to detect organizational differences between the blocked and alternated lists and to modify their processing selectively for the blocked lists. This processing may have emphasized semantic relations, which subsequently inflated false recall. We evaluated this possibility in Experiment 2.

Experiment 2: Between-subjects manipulation of list type

Given the unexpected finding that blocking inflated false recall on an immediate test, we examined whether this blocking effect was due to our presenting blocked lists within the context of alternated lists. As was mentioned in the preceding discussion, one possibility is that exposure to both blocked and alternated list types may have made the thematic differences between the lists more salient. This in turn could cause participants to pay greater attention to the semantic/thematic properties of the list than if they had solely been exposed to blocked or alternated lists, as in Hutchison and Balota (2005). It has been shown in other work that list structure can have a strong effect on the type of encoding processing used at study (Hunt & Einstein, 1981). Specifically, when the relations among list items are obvious, as in a related list, participants are likely to engage in relational processing (Huff & Bodner, 2014), a processing type that has been shown to inflate the DRM illusion (Huff & Bodner, 2013; McCabe, Presmanes, Robertson, & Smith, 2004). Thus, the salient list differences in the within-subjects design may have inflated relational processing for the blocked lists and possibly reduced relational processing for the alternated lists, producing a large difference in false recall between list types.

A simple way to evaluate this possibility would be through the use of a between-subjects design for blocked and alternated lists, since viewing a blocked list would preclude exposure to an alternated list. To be sure, we are not arguing that the use of a between-subjects design results in process-pure encoding of the lists, but instead mitigates any carryover in processing that might occur due to exposure to both list types. In addition, the use of a between-subjects design more closely matched the conditions of Hutchison and Balota (2005), which allowed us to more closely isolate the effects of delay in their design. The predictions therefore remained the same as from Experiment 1: If false memories are governed strictly by associative activation, false recall of the CLs should be equal for the blocked and alternated lists. If false memories are governed by gist-based organization, however, false recall should be greater for blocked than for alternated lists.

Method

Participants A total of 94 individuals were recruited as participants for monetary compensation using Amazon's Mechanical Turk (see Mason & Suri, 2012, for an overview). The participants reported proficiency in English and resided within the United States. Their mean reported age was 37.49 years ($SD = 11.94$, range = 19–68), and their mean education was 15.79 years ($SD = 2.22$, range = 12–22) of formal education.

Design A 2 (Blocking: blocked vs. alternated) \times 2 (Delay: no delay vs. 2.5-min delay) between-subjects design was used.

Materials and procedure All procedures and stimulus lists used were the same as those in Experiment 1, with the following exceptions. First, the filler task in the delay group was changed to a more engaging Tetris task, in an attempt to ensure that online participants were completing it. Second, participants were exposed to lists that were only either blocked or alternated by meaning (see Table 1). List order was counterbalanced across subjects.

Results

The proportions of correct recall and CL false recall, as well as ARC clustering scores, are reported in Table 2. Using the same criterion for calculating ARC scores as in Experiment 1, 96% of the lists were available for analysis in the no-delay group, and 87% of the lists were available for analysis in the delay group. These rates were similar to those from Experiment 1.

Correct recall Unlike in Experiment 1, the effect of blocking was not significant, $F(1, 90) = 1.07$, $MSE = .01$, $p = .30$, though recall was numerically greater for blocked than for alternated lists (.57 vs. .55). An effect of delay was found, $F(1, 90) = 16.05$, $MSE = .01$, $\eta_p^2 = .15$, reflecting greater recall for immediate than for delayed tests (.60 vs. .51). The Blocking \times Delay interaction was not significant, $F < 1$.

ARC analysis For the ARC scores, an effect of blocking was found, $F(1, 90) = 62.69$, $MSE = .06$, $\eta_p^2 = .41$, such that clustering scores were greater on blocked than on alternated lists (.61 vs. .22). The effect of delay was not significant, $F(1, 90) = 1.91$, $MSE = .06$, $p = .17$; however, the Blocking \times Delay interaction was significant, $F(1, 90) = 4.26$, $MSE = .06$, $\eta_p^2 = .05$, and showed the same pattern as in Experiment 1. Specifically, for blocked lists, clustering did not differ between the immediate and delayed tests (.62 vs. .59), $t < 1$, but for alternated lists, clustering increased between the no-delay and delayed tests (.13 vs. .30), $t(44) = 2.05$, $SEM = .06$, $d = 0.62$.

False recall The effects of blocking and delay were not reliable, $F < 1$ and $F(1, 90) = 2.40$, $MSE = .03$, $p = .13$, respectively, but critically, the interaction was significant, $F(1, 90) = 8.46$, $MSE = .03$, $\eta_p^2 = .09$. Follow-up tests revealed that, on an immediate test, the numerically greater false recall for alternated than for blocked lists approached, but did not reach, significance, $t(46) = 1.85$, $SEM = .03$, $p = .07$, $d = 0.54$. However, consistent with our predictions, false recall on the delayed test

was greater for blocked than for alternated lists, $t(44) = 2.23$, $SEM = .04$, $d = 0.67$.

Extralist intrusions per list were found to be quite rare ($M = .23$), and most of these intrusions were related to the meanings presented within the studied list ($M = .18$). No differences in extralist intrusions were found between list types, $F < 1$, though extralist intrusions were marginally greater on delayed than on immediate tests (.29 vs. .17), $F(1, 90) = 3.86$, $MSE = .08$, $p = .06$. The interaction was not significant, $F(1, 90) = 2.38$, $MSE = .08$, $p = .13$.

Discussion

Experiment 2 further evaluated the differences between blocked and alternated list types as a function of immediate versus delayed tests. In a between-subjects design, correct recall was expectedly lower after a delay, but was equivalent between blocked and alternated list types, a finding that is consistent with other encoding contexts, such as generation and production, that have shown diminished or eliminated effects under between-subjects designs (Bertsch, Pesta, Wiscott, & McDaniel, 2007; Fawcett, 2013). Importantly, false recall was equivalent between blocked and alternated lists on an immediate test, a pattern consistent with Hutchison and Balota (2005) and AMT, but was greater on blocked lists after a delay, consistent with the notion that gist-based processes are more influential after a delay.³

ARC clustering scores were calculated to further elucidate these processes. The ARC scores were greater for blocked than for alternated lists. Moreover, clustering increased across a delay, but selectively for alternated lists, a pattern consistent with thematic information producing a greater influence after

³ An anonymous reviewer suggested that the two meanings presented in homograph lists might show different contributions to CL false recall. To explore this possibility, we separated list item recall for the meanings that were presented first in the list (Meaning 1) and second in the list (Meaning 2), to determine whether the order of meanings was related to false recall in Experiment 2. Note that ARC could not be computed for each individual meaning, and therefore only list item recall was available. We then correlated correct recall for each meaning with CL false recall. For blocked lists, we found a significant negative correlation between correct and false recall between Meaning 1 ($r = -.41$, $p < .01$) and Meaning 2 ($r = -.34$, $p < .02$). For alternated lists, the correlations between correct and false recall were similar between meanings, but not significant ($r = -.02$ and $r = .04$, for Meanings 1 and 2, respectively). These relationships were similar on both immediate and delayed tests. Given the significant correlations between meaning and false recall on the blocked lists, we conducted a regression analysis to determine whether Meaning 1 or 2 accounted for unique variance in false recall. The results from the regression analysis showed that Meaning 1 and 2 recall accounted for significant variance in the model [$R^2 = .18$, $F(1, 46) = 9.21$], but Meaning 1 was only a marginally significant predictor ($\beta = -.32$, $p = .07$), and Meaning 2 recall was not significant ($\beta = -.14$, $p = .45$). Thus, it appears that Meaning 1 recall may account for the unique variance on false recall on blocked lists.

a delay.⁴ In contrast, however, clustering did not increase for blocked lists across a delay. Although at first glance this pattern may be considered inconsistent with our argument that gist-based processes increase over a delay, the clustering scores were already quite high, and it is possible that clustering may have been at ceiling and been exempt from any further increase due to the delay. Furthermore, clustering has been shown to be positively related to correct recall (Hunt & Einstein, 1981; Mulligan, 2005). Given that correct recall decreases across a delay for blocked lists, this relationship may further hamper any clustering increase across delays for blocked lists. Thus, a combination of these processes may have contributed to similar clustering for blocked lists on immediate and delayed tests.

General discussion

The purpose of the present experiments was to evaluate the contributions of gist-based versus activation-monitoring processes in the creation of false memories in an associated-list paradigm. To that end, we compared homograph word lists that contained items related to two separate meanings of a single homograph CL that was not studied. Importantly, the lists were organized at study such that the words were either blocked, to enhance the processing of gist-based information, or instead were alternated by meaning. These list types were compared on free recall tests that were completed either immediately after study or after a delay. Supporting past research, correct recall was greater for lists that were blocked by meaning than for those that were alternated (Mather et al., 1997; McDermott, 1996; Toglia et al., 1999), and correct recall of both list types decreased as the delay increased (Payne et al., 1996). Also consistent with past research, by Seamon et al. (2002), and suggestive of a greater reliance on gist-based memory information over time, CL false recall did not decrease over the retention interval.

In addition to the overall pattern of false recall, the clustering scores also suggest increased reliance on gist over a delay, especially in alternated lists. Although the availability of gist-based information was relatively low for alternated lists,

⁴ We were also interested in whether ARC scores were related to CL false recall, given that clustering is likely related to the use of gist information. ARC was not related to false recall on alternated lists ($r = -.07, p = .69$), but it was related to false recall on blocked lists ($r = -.29, p < .05$). Given that ARC is often positively related to correct recall and correct recall is often negatively related to CL false recall (Roediger et al. 2001a, b), we entered ARC scores and correct recall into a regression analysis to determine whether the negative relationship between false recall and ARC on blocked lists held when correct recall was entered into the model. Both ARC scores and correct recall accounted for significant variance in the model [$R^2 = .21, F(1, 46) = 6.13$]; however, correct recall was a significant predictor ($\beta = -.37, p < .01$), whereas ARC was not ($\beta = -.21, p = .12$). Thus, the negative relationship between CL false recall and ARC is likely due to clustering increasing correct recall, though this relationship was not quite significant in Experiment 2 ($r = .21, p = .16$, for the correlation between ARC and correct recall).

participants still actively utilized thematic information to guide their recall, an observation supported by the positive ARC scores for alternated lists following the delay. This pattern, with an increase in clustering over the delay for the alternated lists, is difficult to reconcile with the AMT account, and is instead consistent with a gist extraction account. ARC scores confirmed that clustering significantly increased after a delay for the alternated lists, but it numerically decreased for blocked lists. This crossover of clustering effects suggests that people rely less on serial-order information to guide recall following a delay (Howard & Kahana, 1999), which coincides with FTT, such that verbatim traces fade and gist traces persist over time (Brainerd & Reyna, 2002). If participants were relying on strict serial order, ARC scores should have polarized closer to -1.0 for alternated lists and $+1.0$ for blocked lists, and these polarized clustering scores would be expected to gravitate toward zero over a delay. Although such a simple reduction in reliance on serial order following delay is consistent with a crossover interaction, the alternated recall pattern does not fit this explanation. Instead, for alternated lists, ARC scores were near zero immediately, but increased significantly above zero after a delay, indicating active organization by meaning over time.

Unlike the similar patterns across Experiments 1 and 2 for veridical recall and clustering scores, the patterns for false recall were quite different across experiments. In Experiment 1, blocking inflated false recall on an immediate test, an effect that disappeared following a delay. Although unanticipated, the results were consistent with the clustering scores in suggesting that the Experiment 1 participants initially engaged in less gist-related processing for alternated than for blocked lists, but engaged in more active reorganization of alternated lists to guide recall after the delay. The increase in gist-related processing after the delay eliminated any false recall differences across lists. Experiment 2 used a between-subjects manipulation akin to that of Hutchison and Balota (2005), to examine the possibility that the unanticipated pattern may have been due to the use of a within-subjects manipulation and carryover effects that made list differences more salient.⁵ Indeed, our obtained pattern was in line with our original prediction that an advantage of blocking list items during study would only emerge following a delay, at which time reliance on gist information should increase.

As we summarized in the introduction, false memory effects have been found following delays in excess of one week (e.g., Seamon et al., 2002). Our 2.5-min delay was considerably shorter than these longer retention intervals; however, we note that even with our shorter retention interval, a reduction in verbatim details was evidenced by a reduction in correct

⁵ Prior to conducting Experiment 2, we first examined just the initial two lists from Experiment 1, which were alternated for half of the participants and blocked for the other half of the participants. Consistent with Hutchison and Balota's (2005) results and Experiment 2, the blocking advantage on immediate recall disappeared in this between-subjects analysis of just the initial lists.

recall. It is unclear how verbatim and gist traces differ between relatively short delays such as ours and much longer delays; however, the delay used in our experiments was likely sufficient to demonstrate a loss of verbatim details. Importantly, the reduction in verbatim traces across the delay was not so great as to reduce correct recall enough to preclude the calculation of ARC scores. Thus, the delay used in our experiments was sufficient to allow for a measurement of verbatim and gist traces at both immediate and delayed retention intervals.

It is important to note, however, that although the present results clearly demonstrate the importance of gist processing to false memory, they also do not undermine Hutchison and Balota's (2005) main findings that supported AMT. Specifically, Hutchison and Balota's main evidence for AMT, as opposed to FTT, was that, across five experiments, the increases in false memory obtained when adding six related items to a list were the same, regardless of whether those items were of the same meaning or a different meaning, and this occurred despite participants reporting greater relatedness when the additional items were of the same meaning (Exp. 6). Such results are inconsistent with FTT, and instead argue for the importance of lexical associative activation in the formation of false memories. Similarly, the equal (in fact, marginally greater) immediate false recall for alternated lists in Experiment 2 replicates Hutchison and Balota's study and demonstrates the importance of associative activation for false memory. However, on the basis of our other findings in the present study, we argue that gist-based processes become increasingly important in producing false memory after a delay.

Through our present design, we also sought to improve upon a potential flaw in Hutchison and Balota's (2005) lists, in which the BAS ordering within the study lists was not presented equally between blocked and alternated list types. As was summarized in the introduction, we structured the alternated lists to reflect a scalloped presentation of BASs, as in the blocked list types (see, too, Table 1). In Experiment 2, our immediate-testing condition produced correct and false recall patterns similar to those reported by Hutchison and Balota, suggesting that differences between blocked and alternated lists types were not due to a confound in BAS presentation. One departure, however, was a marginal difference in false recall between the blocked and alternated lists, in which alternated lists produced numerically greater false recall than did blocked lists. This pattern is interesting, because neither AMT nor FTT predict that false recall would be greater for alternated than for blocked lists. We emphasize caution in overinterpreting a marginal difference, but note one speculation that may account for this pattern: Specifically, in the present lists the newly normed items were arranged into correct BAS order, whereas the lists used by Hutchison and Balota always had the normed items in the latter serial positions. It is therefore possible that descending BAS ordering is important for increasing false recall on alternated lists.

Finally, it is worth mentioning that the participants in Experiments 1 and 2 were taken from separate subject pools. The participants in Experiment 1 were university undergraduates, likely between the ages of 18 and 22, whereas the Mechanical Turk participants in Experiment 2 were likely considerably older and more highly educated, due to their increased age. Sample differences could possibly account for the different patterns found in our two experiments, aside from the within-/between-list type manipulations used. However, the immediate test group in Experiment 2 showed similar results to Hutchison and Balota, who utilized an undergraduate sample. The similarities in these results give us confidence that the differences across our experiments are not simply due to sample differences. Furthermore, we also argue that the relatively more diverse Mechanical Turk sample is also likely more representative of the greater population, which improves the external validity of our reported results (Mason & Suri, 2012).

Conclusions

The present study was designed to test the extent to which gist-based processing contributes to false memory, especially following a delay. The results of this study were noteworthy in that both experiments showed an increase in CL false recall due to blocking. However, the blocking increase depended on the use of a within- or between-subjects design. Luckily, the clustering analyses were more straightforward, such that clustering increased for alternated lists across the delay for both experiments, suggesting greater use of gist-based processing to guide retrieval. This use of gist contributes to false memory in two ways. First, blocking lists provides a theme-based organizational framework to guide encoding during study. Phrased differently, blocking lists could lead to enhanced relational processing during study, which has been shown to increase associative false memory (Huff & Bodner, 2013) and to produce greater clustering in free recall (Hunt & Einstein, 1981). Second, in addition to the material-induced organization provided by blocked lists, active gist-guided retrieval during recall increases the likelihood that the CL will be produced (Payne et al., 1996). In addition to gist, this study has also demonstrated the importance of associative activation in contributing to the creation of false memory, at least during immediate testing, at which point retrieval is less active and effortful. Therefore, the present data add to previous studies that have shown the importance of associative activation, but they also provide strong evidence for the importance of gist-based processing to the creation of false memories.

Author note Special thanks to Kelly Buchanan, Reema Najjar, and Hailey Little for running participants and scoring the recall data for this study. Funding support was provided by NIH Grant Number T32 AG000030-32 to the first author.

Appendix 1

Table 3 Alternated homograph list items and BAS values for the first set of 7 items (Sublist 1) and the second set of 7 items (Sublist 2)

<i>“Bat” List Total BAS = 1.08</i>						
Sublist 1:	Baseball	Blind	Softball	Fangs	Club	Rabies*
BAS	.31	.13	.09	.01	.02	.00
Sublist 2:	Ball	Vampire	Hit	Cave	Glove	Cavern
BAS	.19	.21	.04	.05	.02	.01
<i>“Bow” List Total BAS = 1.23</i>						
Sublist 1:	Arrow	Tie	Indian	Gift	Cross	Decoration
BAS	.53	.10	.03	.01	.02	.00
Sublist 2:	Archery*	Ribbon	Violin	Sash	Hunter*	Present
BAS	.24	.26	.03	.01	.00	.00
<i>“Cell” List Total BAS = 1.03</i>						
Sublist 1:	Membrane*	Prison	Nerve	Inmate	Bacteria	Cage
BAS	.48	.10	.10	.05	.02	.01
Sublist 2:	Nucleus	Holding	Molecule	Jail	Germ	Bars*
BAS	.10	.16	.04	.07	.01	.04
<i>“Class” List Total BAS = 0.95</i>						
Sublist 1:	Course	Group	Lecture	Status	Attendance	Social
BAS	.19	.05	.12	.01	.10	.00
Sublist 2:	Subject	Stereotype	Lesson	Economic	Professor	Prestige
BAS	.14	.20	.11	.02	.00	.01
<i>“Fall” List Total BAS = 3.20</i>						
Sublist 1:	Stumble	Season	Rise	Leaves	Faint	Harvest
BAS	.70	.19	.33	.17	.21	.00
Sublist 2:	Slip	Autumn	Trip	Spring	Clumsy	Brisk
BAS	.51	.52	.29	.19	.09	.00
<i>“Fly” List Total BAS = 3.10</i>						
Sublist 1:	Swatter	Soar	Moth	Birds	Bug	Glide
BAS	.74	.37	.10	.32	.06	.10
Sublist 2:	Maggot	Kite	Insect	Airplane	Mosquito	Wings
BAS	.15	.57	.09	.32	.05	.23
<i>“Foot” List Total BAS = 2.55</i>						
Sublist 1:	Toe	Yard	Shoe	Meter	Leg	Length
BAS	.60	.12	.32	.02	.20	.00
Sublist 2:	Ankle	Inch	Heel	Mile	Sock	Measurement
BAS	.36	.47	.22	.05	.17	.02
<i>“Head” List Total BAS = 2.11</i>						
Sublist 1:	Chairperson	Hat	Leadership	Scalp	Boss	Hair
BAS	.13	.35	.04	.32	.03	.04
Sublist 2:	Chief	Skull	Department	Neck	Principal	Brain
BAS	.05	.49	.04	.33	.02	.27
<i>“Horn” List Total BAS = 1.85</i>						
Sublist 1:	Honk	Unicorn	Trumpet	Bull	Trombone	Antelope
BAS	.52	.11	.24	.05	.06	.03
Sublist 2:	Bugle	Rhinoceros*	Tuba	Ram	Brass	Antler
BAS	.30	.24	.08	.11	.04	.05
<i>“Iron” List Total BAS = 1.13</i>						
Sublist 1:	Crease	Steel	Steam	Rust	Wrinkle	Element

Table 3 (continued)

BAS	.09	.22	.09	.05	.06	.02
Sublist 2:	Starch	Ore	Press	Metal	Shirt	Copper
BAS	.09	.32	.06	.10	.00	.03
<i>“Letter” List Total BAS = 2.39</i>						
Sublist 1:	Scarlet*	Stamp	Greek	Mail	Symbol	Recommendation
BAS	.48	.38	.04	.24	.02	.11
Sublist 2:	Number	Envelope	Bold	Note	Print	Send
BAS	.17	.49	.02	.29	.01	.15
<i>“Race” List Total BAS = 1.23</i>						
Sublist 1:	Drag	Minority	Track	Prejudice	Winner	Culture
BAS	.21	.06	.10	.04	.04	.01
Sublist 2:	Compete	Ethnicity*	Runner	Bias	Horse	Stereotype
BAS	.12	.48	.05	.06	.03	.04
<i>“Right” List Total BAS = 2.81</i>						
Sublist 1:	Wrong	Handed*	Accurate	Clockwise	Exact	Direction
BAS	.72	.32	.16	.06	.07	.05
Sublist 2:	Correct	Left	Proper	Starboard	Answer	Turn
BAS	.23	.93	.07	.10	.04	.06
<i>“Watch” List Total BAS = 1.65</i>						
Sublist 1:	Observe	Digital	Guard	Clock	Look	Tick
BAS	.50	.20	.11	.08	.03	.01
Sublist 2:	Television	Wrist*	Neighborhood	Time	Patrol	Pocket
BAS	.12	.34	.08	.15	.03	.04

* Cue words not found in Nelson et al. (1999) that were not included in the words lists according to BAS strength.

Appendix 2

Table 4 Blocked homograph list items and BAS values for the first set of 7 items (Sublist 1) and the second set of 7 items (Sublist 2)

<i>“Bat” List Total BAS = 1.08</i>						
Sublist 1:	Baseball	Ball	Softball	Hit	Club	Glove
BAS	.31	.19	.09	.04	.02	.02
Sublist 2:	Vampire	Blind	Cave	Fangs	Cavern	Rabies*
BAS	.21	.31	.05	.01	.01	.00
<i>“Bow” List Total BAS = 1.23</i>						
Sublist 1:	Arrow	Archery*	Indian	Violin	Cross	Hunter*
BAS	.53	.24	.03	.03	.02	.00
Sublist 2:	Ribbon	Tie	Sash	Gift	Present	Decoration
BAS	.26	.10	.01	.01	.00	.00
<i>“Cell” List Total BAS = 1.03</i>						
Sublist 1:	Membrane*	Nucleus	Nerve	Molecule	Bacteria	Germ
BAS	.48	.10	.10	.04	.02	.01
Sublist 2:	Holding	Prison	Jail	Inmate	Bars*	Cage
BAS	.16	.10	.07	.05	.04	.01
<i>“Class” List Total BAS = 0.95</i>						
Sublist 1:	Course	Subject	Lecture	Lesson	Attendance	Professor
BAS	.19	.14	.12	.11	.10	.00
Sublist 2:	Stereotype	Group	Economic	Status	Prestige	Social
BAS	.20	.05	.02	.01	.01	.00
<i>“Fall” List Total BAS = 3.20</i>						
Sublist 1:	Stumble	Slip	Rise	Trip	Faint	Clumsy
BAS	.70	.51	.33	.29	.21	.09
Sublist 2:	Autumn	Season	Spring	Leaves	Brisk	Harvest
BAS	.52	.19	.19	.17	.00	.00

Table 4 (continued)

“Fly” List Total BAS = 3.10						
Sublist 1:	Swatter	Maggot	Moth	Insect	Bug	Mosquito
BAS	.74	.15	.10	.09	.06	.05
Sublist 2:	Kite	Soar	Airplane	Birds	Wings	Glide
BAS	.57	.37	.32	.32	.23	.10
“Foot” List Total BAS = 2.55						
Sublist 1:	Toe	Ankle	Shoe	Heel	Leg	Sock
BAS	.60	.36	.32	.22	.20	.17
Sublist 2:	Inch	Yard	Mile	Meter	Measurement	Length
BAS	.47	.12	.05	.02	.02	.00
“Head” List Total BAS = 2.11						
Sublist 1:	Chairperson	Chief	Leadership	Department	Boss	Principal
BAS	.13	.05	.04	.04	.03	.02
Sublist 2:	Skull	Hat	Neck	Scalp	Brain	Hair
BAS	.49	.35	.33	.32	.27	.04
“Horn” List Total BAS = 1.85						
Sublist 1:	Honk	Bugle	Trumpet	Tuba	Trombone	Brass
BAS	.52	.30	.24	.08	.06	.04
Sublist 2:	Rhinoceros*	Unicorn	Ram	Bull	Antler	Antelope
BAS	.24	.11	.11	.05	.05	.03
“Iron” List Total BAS = 1.13						
Sublist 1:	Crease	Starch	Steam	Press	Wrinkle	Shirt
BAS	.09	.09	.09	.06	.06	.00
Sublist 2:	Ore	Steel	Metal	Rust	Copper	Element
BAS	.32	.22	.10	.05	.03	.02
“Letter” List Total BAS = 2.39						
Sublist 1:	Scarlet*	Number	Greek	Bold	Symbol	Print
BAS	.48	.17	.04	.02	.02	.01
Sublist 2:	Envelope	Stamp	Note	Mail	Send	Recommendation
BAS	.49	.38	.29	.24	.15	.11
“Race” List Total BAS = 1.23						
Sublist 1:	Drag	Compete	Track	Runner	Winner	Horse
BAS	.21	.12	.10	.05	.04	.03
Sublist 2:	Ethnicity*	Minority	Bias	Prejudice	Stereotype	Culture
BAS	.48	.06	.06	.04	.04	.01
“Right” List Total BAS = 2.81						
Sublist 1:	Wrong	Correct	Accurate	Proper	Exact	Answer
BAS	.72	.23	.16	.07	.07	.04
Sublist 2:	Left	Handed*	Starboard	Clockwise	Turn	Direction
BAS	.93	.32	.10	.06	.06	.05
“Watch” List Total BAS = 1.65						
Sublist 1:	Observe	Television	Guard	Neighborhood	Look	Patrol
BAS	.50	.12	.11	.08	.03	.03
Sublist 2:	Wrist*	Digital	Time	Clock	Pocket	Tick
BAS	.34	.20	.15	.08	.04	.01

* Cue words not found in Nelson et al. (1999) that were not included in the words lists according to BAS strength.

References

- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22, 261–295. doi:10.1016/S0022-5371(83)90201-3
- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, 79, 97–123. doi:10.1037/h0033773
- Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 827–845. doi:10.1037/0278-7393.22.4.827
- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35, 201–210.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49, 229–240.
- Brainerd, C. J., Payne, D. G., Wright, R., & Reyna, V. F. (2003). Phantom recall. *Journal of Memory and Language*, 48, 445–467. doi:10.1016/S0749-596X(02)00501-6
- Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, 10, 3–47.
- 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., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychonomic Bulletin & Review*, 15, 1035–1053. doi:10.3758/PBR.15.6.1035

- Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 117–156). Mahwah, NJ: Erlbaum.
- Cann, D. R., McRae, K., & Katz, A. N. (2011). False recall in the Deese–Roediger–McDermott paradigm: The roles of gist and associative strength. *Quarterly Journal of Experimental Psychology*, *64*, 1515–1542. doi:10.1080/17470218.2011.560272
- Cofer, C. N. (1967). Does conceptual organization influence the amount retrained in immediate free recall? In B. Kleinmuntz (Ed.), *Concepts and the Structure of Memory*. New York: Wiley.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428. doi: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*, 17–22. doi:10.1037/h0046671
- Fawcett, J. M. (2013). The production effect benefits performance in between-subjects designs: A meta-analysis. *Acta Psychologica*, *142*, 1–5. doi:10.1016/j.actpsy.2012.10.001
- Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, *4*, 271–276. doi:10.3758/BF03209405
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 923–941. doi:10.1037/0278-7393.25.4.923
- Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1246–1256. doi:10.1037/a0031338
- Huff, M. J., & Bodner, G. E. (2014). All varieties of encoding variability are not created equal: Separating variable processing from variable tasks. *Journal of Memory and Language*, *73*, 43–58.
- Huff, M. J., Bodner, G. E., & Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: A meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psychonomic Bulletin & Review*, *22*, 349–365. doi:10.3758/s13423-014-0648-8
- Huff, M. J., Coane, J. H., Hutchison, K. A., Grasser, E. B., & Blais, J. E. (2012). Interpolated task effects on direct and mediated false recognition: Effects of initial recall, recognition, and the ironic effect of guessing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1720–1730. doi:10.1037/a0028476
- Huff, M. J., & Hutchison, K. A. (2011). The effects of mediated word lists on false recall and recognition. *Memory & Cognition*, *39*, 941–953. doi:10.3758/s13421-011-0077-0
- Huff, M. J., Meade, M. L., & Hutchison, K. A. (2011). Age-related differences in guessing on free and forced recall tests. *Memory*, *19*, 317–330.
- Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*, 497–514.
- Hunt, R. R., Smith, R. E., & Dunlap, K. R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language*, *65*, 378–389.
- 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*, 1–28. doi:10.1016/j.jml.2004.08.003
- Hutchison, K. A., Meade, M. L., Williams, N., McNabb, J., & Manley, K. (2015). Preceding context and initial orthographic similarity interact to increase false recollection in the DRM paradigm. Manuscript in preparation.
- Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*, 577–581. doi:10.3758/BF03214352
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. doi:10.1037/0033-2909.114.1.3
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review*, *6*, 130–135. doi:10.3758/BF03210820
- Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 328–372). New York, NY: Academic Press. Reprinted in G. H. Bower (Ed.), *Human memory: Basic processes*. New York, NY: Academic Press, 1977.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*, 1–23. doi:10.3758/s13428-011-0124-6
- Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 3–23. doi:10.1037/0278-7393.21.1.3
- Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*, 826–837.
- McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review*, *11*, 1074–1079. doi:10.3758/BF03196739
- McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, *30*, 1065–1077.
- McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language*, *35*, 212–230. doi:10.1006/jmla.1996.0012
- McDermott, K. B., & Roediger, H. L., III. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language*, *39*, 508–520. doi:10.1006/jmla.1998.2582
- McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What is the connection between true and false memories? the differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1177–1194.
- Meade, M. L., Hutchison, K. A., & Rand, K. (2010). Effects of delay and number of related list items on implicit activation for DRM critical items in a speeded naming task. *Journal of Memory and Language*, *62*, 302–310.
- Meade, M. L., Watson, J. M., Balota, D. A., & Roediger, H. L., III. (2007). The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. *Journal of Memory and Language*, *56*, 305–320. doi:10.1016/j.jml.2006.07.007
- Miller, M. B., Guerin, S. A., & Wolford, G. L. (2011). The strategic nature of false recognition in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1228–1235. doi:10.1037/a0024539
- Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, *106*, 398–405. doi:10.1037/0033-295X.106.2.398
- Mulligan, N. W. (2005). Total retrieval time and hypermnnesia: Investigating the benefits of multiple recall tests. *Psychological Research*, *69*, 272–284. doi:10.1007/s00426-004-0178-5

- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). The University of South Florida word association, rhyme, and word fragment norms [Database]. Available from <http://w3.usf.edu/FreeAssociation/>
- 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*, 261–285.
- Rabinowitz, J. C., Craik, F. I., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology*, *36*, 325–344.
- Reyna, V. F. (1998). Fuzzy-trace theory and false memory. In M. Intons-Peterson & D. Best (Eds.), *Challenges and controversies: Memory distortions and their prevention* (pp. 15–27). Mahwah, NJ: Erlbaum.
- Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001a). 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.
- 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. (2001b). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385–407. doi:10.3758/BF03196177
- Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*, 45–48. doi:10.1037/h0031355
- Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition*, *30*, 1054–1064. doi:10.3758/BF03194323
- Senkova, O., & Otani, H. (2012). Category clustering calculator for free recall. *Advances in Cognitive Psychology*, *8*, 292–295. doi:10.5709/acp-0124-y
- 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
- Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.
- Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, *7*, 233–256.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Oxford University Press.
- Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122–129. doi:10.1037/h0022014
- Zaromb, F. M., & Roediger, H. L., III. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, *38*, 995–1008. doi:10.3758/MC.38.8.995