Testing potentiates new learning in the misinformation paradigm

Leamarie T. Gordon & Ayanna K. Thomas

Memory & Cognition

ISSN 0090-502X

Mem Cogn
DOI 10.3758/s13421-013-0361-2

Mem Cogn
DOI 10.3758/s13421-013-0361-2

Testing potentiates new learning in the misinformation paradigm

Leamarie T. Gordon · Ayanna K. Thomas

Abstract Retrieval enhanced suggestibility (RES) is the finding that the misinformation effect is exacerbated when a test precedes misleading postevent information (Chan, Thomas, & Bulevich Psychological Science 20: 66–73, 2009). In the present study, we tested three hypotheses relevant to RES. First, we examined whether retrieval of critical details was necessary for the RES effect. Second, we examined whether initial testing influenced the allocation of attention to critical details during postevent information processing. Finally, we examined whether RES resulted in impaired access to the originally learned information. We compared three groups of participants in three experiments: an identical-test group, a related-test group, and a standard misinformation group. Both testing groups were tested on the original event before the introduction of misinformation; however, the identical-test group took the same test before and after the misinformation, whereas the related-test group took different tests before and after misinformation. We found that testing before misleading postevent information affected attention allocation to details in the postevent narrative. Furthermore, the RES effect did not accompany reduced accessibility to the original information, as measured by a modified–modified free recall test. These data have implications for how testing may potentiate new learning.

Keywords Eyewitness memory · Misinformation effect · New learning · Retrieval · Testing effect

After witnessing a crime, most individuals undergo some form of immediate recall of the event. That is, witnesses are likely to call the police, be interviewed by responders to the scene, or even contact family or friends to talk about what they witnessed. In the present study, we investigated the role of immediate testing on eyewitness memory in the misinformation paradigm. Numerous studies have demonstrated the deleterious effects that misleading postevent information can have on memory for an originally witnessed event (see Loftus, 2005, for a review). More recent studies have suggested that in certain situations, immediate retrieval may exacerbate misinformation susceptibility (e.g., Chan et al., 2009), whereas in other situations, immediate retrieval has been shown to reduce misinformation susceptibility (Gabbert, Hope, Fisher, & Jamieson, 2012). The circumstances that result in enhanced or reduced misinformation susceptibility due to immediate testing remain unclear. In the present study, we tested the hypothesis that immediate testing directly impacts posttest learning, resulting in enhanced posttest learning or misinformation susceptibility. We further hypothesized that the seemingly deleterious effects of immediate testing may be diminished through unconstrained retrieval.

The hypothesis that testing after witnessing an event will reduce susceptibility to misleading postevent information has emerged from the well-established testing effect. Research has demonstrated that repeated testing increases long-term retention, as compared to repeated study (for a review, see Roediger & Karpicke, 2006). One theory to explain the testing effect is that testing appears to produce an elaboration of an existing memory representation, perhaps in terms of retrieval routes, which offers a more robust memory trace than does simple restudy (Bjork, 1975; Carpenter, 2009, 2011; McDaniel, Kowitz, & Dunay, 1989; Pyc & Rawson, 2012). Alternatively, repeated testing may enhance memory by improving the diagnostic value of retrieval cues (Karpicke & Blunt, 2011; Karpicke & Smith, 2012). Regardless of the underlying mechanism, repeated testing has consistently been shown to improve memory.

L. T. Gordon (✉) · A. K. Thomas
Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155, USA
e-mail: leamarie.gordon@tufts.edu

Published online: 12 September 2013

Though testing effects are robust, few studies have examined the benefits of repeated testing in eyewitness scenarios, and those that do exist have yielded conflicting results. On the one hand, a series of studies by Chan and colleagues (Chan & Langley, 2011; Chan & LaPaglia, 2011; Chan et al., 2009; Thomas, Bulevich, & Chan, 2010) have demonstrated that testing before misleading postevent information increased susceptibility to misinformation. Comparing a group who took a test before misleading postevent information to a standard misinformation group on a final test of memory, Chan and colleagues found that immediate testing impaired accuracy on the final test of memory and increased misinformation production. The retrieval enhanced suggestibility (RES) effect has been replicated in several subsequent studies (Chan & Langley, 2011; Chan & LaPaglia, 2011; Thomas et al., 2010).

On the other hand, research has demonstrated that immediate retrieval of a witnessed event may immunize participants against the deleterious effects of misleading postevent information (Gabbert et al., 2012; Lane, Mather, Villa, & Morita, 2001; Memon, Zaragoza, Clifford, & Kidd, 2010; Saunders & MacLeod, 2002; Thomas et al., 2010). In one study, Lane et al. had participants provide a review of a witnessed event prior to the introduction of misinformation. They found that there was no difference in misinformation susceptibility between the review condition and a no-review control condition. In other words, testing prior to misinformation did not increase susceptibility. In a related study, Gabbert et al. (2012) had participants view a video of a robbery, and then complete a self-administered interview (SAI; Gabbert, Hope, & Fisher, 2009). They found that participants who had completed a SAI were less susceptible to misinformation than those who did not (Gabbert et al., 2012). In both of these studies, participants had more control over their initial retrieval than in the studies performed by Chan and colleagues. Specifically, in Lane et al. (2001) participants were given free recall instructions. The SAI employed by Gabbert and colleagues offered a more structured recall process than free recall; however, specific details in the original event were not directly queried. In standard RES studies, participants are tested on specific details, and those details are then manipulated in the context of a postevent narrative. Therefore, in the context of the RES methodology, immediate retrieval is tied to later misinformation presentation. In the context of less constrained initial-testing studies, retrieval of details that will later be manipulated is not guaranteed.

The conflict in the literature suggests that the nature of the initial test may determine misinformation susceptibility. The more closely that initial testing relates to the posttest narrative, the more likely a RES effect. In three experiments, we examined this relationship by manipulating the level of association between initial testing and the posttest narrative. We tested the hypothesis that retrieval, or attempted retrieval, of details later manipulated in the narrative results in RES, whereas retrieval of details associated with the original event that are not later manipulated in a postevent narrative may reduce RES and enhanced memory for the original event. We compared two kinds of initial tests. The first was analogous to tests used in previous RES experiments. In this testing group, participants retrieved or attempted to retrieve details that were later manipulated in the narrative. This group was compared to an initial-testing group that retrieved or attempted to retrieve details associated with the original event that were *not* later manipulated in the narrative. We hypothesized that initial or immediate testing would affect misinformation susceptibility, but that the magnitude of susceptibility would be contingent on the retrieval or attempted retrieval of details later manipulated in the narrative.

Experiment 1

The goal of Experiment 1 was to test the hypothesis that retrieval of critical details, prior to exposure to misleading information about those same details, is necessary for RES to occur. In addition to the single-test control and identical-test groups from standard RES studies, we created a second, repeated-test group in which initial- and final-test questions were contextually related, but not identical. We hypothesized that if retrieval of critical details is necessary for RES to occur, the related-test group would not show increased susceptibility to misinformation as compared to the single-test control group. The identical-test group, on the other hand, would show the standard RES effect, demonstrated by both an increase in misinformation production and a decrease in accuracy on misleading trials as compared to the single-test control group.

Method

Participants A group of 89 undergraduate students from Tufts University participated in this study for course research credit.

Materials and procedure The experimental design was a 3 (testing: related, identical, or single) × 3 (item type: consistent, control, or misleading) mixed design. Presence/type of the initial test was manipulated between subjects, whereas item type was manipulated within subjects.

As in Chan et al. (2009), we used the first episode of the first season of the television program *24* as the witnessed event material. The visual narrative was created by modifying the episode guide provided by Fox television at www.fox.com/24. No participant had seen this video before.

The test materials were 48 questions that asked about details from the witnessed event. Questions were divided into 24 pairs denoted as Set A and Set B. Two sets of questions were required for the related-test group. For example, one question in Set A was, "What type of plane does Tony report has exploded?" and the related question in Set B was, "Tony reports that a plane exploded over a desert. Which desert?" Related questions were constructed on the basis of the degree of co-occurrence of queried detail in the witnessed event. That is, we strove to construct them in such a way that answering a question from Set A activated information relevant to the related question from Set B. Crucially, critical details from the video that served as the answers to each question of a pair across Sets A and B were not the same. As per Chan, McDermott, and Roediger (2006), we provided a metric of question-pair relatedness via the pairwise comparison feature on the latent semantic analysis website (http://lsa.colorado.edu). We computed relatedness ratings for both intended pairs, and random pairs. The purpose of this was to show that the intended pairs were associated via *specific* information from the witnessed event. We did so by demonstrating the average relatedness rating for intended pairs exceeded that of the random pair. Indeed, the average relatedness rating for intended pairs (a correlation) was .62, significantly higher than the average of .11 for random pairs, $t(46) = 11.86$, $p < .01$, $d = 3.49$.

Participants first viewed the episode of *24*. After viewing the witnessed event, those in the identical- and related-testing groups took an immediate recall test on 24 details of the video. For the related-test takers, these questions belonged to Set A. For the identical-test takers, these questions belonged to Set B. During the cued-recall test, participants were told to answer every question (by typing their responses into the computer). No corrective feedback was provided, and participants were required to answer every question. Participants in the single-testing group played Tetris (a computerized falling-shape game) for the same amount of time (12 min). Participants in each condition then completed a brief demographic questionnaire and a synonym and antonym vocabulary test (Salthouse, 1993). This distractor phase lasted approximately 20 min.

Following the distractor tasks, participants read a narrative that described the video. The narrative was presented via computer and divided across five screen captures. Each screen presented between three and five paragraphs of the narrative, and all screens included control, consistent, and misleading details. Participants pressed a computer key to advance to the next screen and were unable to return to a previous screen. Twenty-four critical details were targeted in the narrative. The critical details were details that had been viewed in the original event and then correctly repeated, omitted, or distorted in the postevent narrative. Details were randomly assigned to one of the three counterbalancing groups, and each group of eight details appeared equally often as control, consistent, or misleading details. The final test required that participants answer questions that directly assessed memory for these critical details. The final-test questions for both repeated-test groups were always from Set B. Before beginning the final test, participants were instructed to answer every question by reporting what they recalled from the video.

The key manipulation was that, for the identical-test participants, the initial and final tests were the same. Thus, the narrative content pertained directly to *both* tests. However, for the related-test participants, the critical details in the narrative were *related* to the questions from the initial test (for these participants, the Set A questions), but participants were not asked to retrieve information on the initial test that would later be manipulated in the narrative. For example, one initial-test question given to related-test takers asked, "How many joints did Teri find in Kimberly's room?" On the final test, participants were asked "Where in Kimberly's room did Teri find the joints?" When this question pair was presented as part of the consistent condition, the narrative indicated that Teri found the joints in Kimberly's desk. This information was consistent with what participants had seen in the video. When this question pair was presented in the *misleading* condition, the narrative indicated that Teri found the joints in Kimberly's dresser. When the item was in the *control* condition, the narrative indicated only that Teri found joints in Kimberly's "room." In each testing condition, when misinformation was presented, it was always a plausible alternative, and critical information in the narrative was only presented once. Each final-test item was counterbalanced, such that all critical details served equally across participants in the control, consistent, and misleading conditions.

Results

Unless otherwise noted, all results reported were significant at an alpha level of .05. Figure 1 displays the mean probabilities of responding with accurate information from the video or misinformation from the narrative, as a function of test condition and item type. Figure 1 illustrates that test condition affected responding on misleading trials. The three-way interaction of item type (misleading, consistent, control), testing group (single, repeated, related), and response (video, narrative) was significant, $F(4, 172) = 4.38$, $\eta_p^2 = .09$. Planned comparisons using a Bonferroni correction revealed that participants in the identical-testing group ($M = .50$) were marginally less likely to produce correct information than were participants in the related-testing group ($M = .61$), $t(57) = 1.73$, $p = .06$, and significantly less likely to produce correct information than were participants in the single-testing group ($M = .65$), $t(58) = 2.42$, $d = 0.62$. We found no difference in correct responding between the related- and single-test groups. Participants in the identical-test group ($M = .37$) were also more likely to produce misinformation on misleading trials than were those in the single-test group ($M = .19$), $t(58) = 2.9$, $d = 0.75$. Similarly, participants in the related-test group ($M = .32$) were more likely to produce misinformation than were those in the single-test group, $t(57) = 2.52$, $d = 0.64$. No difference in misinformation production was apparent between participants in the related- and identical-test groups.

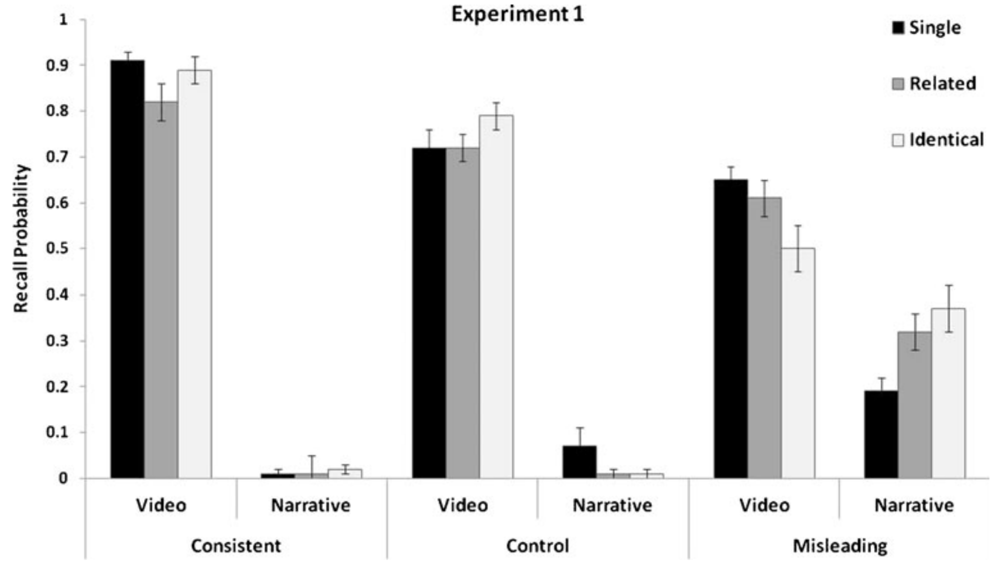

Fig. 1 Mean probabilities of responding with accurate video information or with misinformation from the narrative, as a function of test group and item type in Experiment 1

Discussion

The results of Experiment 1 suggest that retrieval of critical details, or details manipulated in a postevent narrative, may not be necessary for RES. That is, when susceptibility to misinformation was measured as production of misinformation, both the identical- and related-test groups were more likely to produce misinformation than were single-test participants. Although participants in the related-test group took a test about the witnessed event before reading the postevent narrative, they were not questioned about critical details manipulated in the narrative. Thus, critical detail retrieval was not a requirement for RES. Rather, our data suggest that testing before postevent information may direct attention allocation to specific details in the narrative, enhancing learning of that information.

This conclusion is consistent with findings in the verbal-learning literature that have demonstrated the potentiation of new learning after testing. Specifically, Tulving and Watkins (1974) demonstrated greater learning of A–C in an A–B, A–C paired-associate learning paradigm if a test of A–B preceded A–C learning, as compared to when no test intervened between the learning phases. The finding that participants produced more misinformation details in the repeated-testing than in the single-testing groups suggests that immediate testing may result in RES, because it affects learning of the postevent narrative.

Experiment 2

The goal of Experiment 2 was to investigate the influence of immediate testing on posttest processing. The results from the first experiment, as well as those from the verbal-learning domain, (e.g., Tulving & Watkins, 1974) suggest that cued-recall testing prior to new learning will facilitate posttest learning. In Experiment 2, we tested whether an attention mechanism could account for enhanced misinformation production. Specifically, we hypothesized that testing prior to new learning would influence the allocation of attention to details in the narrative. Thomas et al. (2010) suggested that immediate testing in a misinformation paradigm may result in attentional capture by details in a posttest narrative that are related to information previously tested. Thomas et al. hypothesized that attentional capture influenced both the learning of misinformation and the ease, or fluency, with which information from the narrative came to mind during the final test phase (i.e., Baddeley, 1982; Jacoby & Dallas, 1981; Kelley & Lindsay, 1993; Koriat, 1993).

Experiment 1 of the present manuscript provides additional support for the conclusion that immediate testing may influence how posttest information is learned. Experiment 2 was designed to directly measure attention during postevent narrative processing. In this experiment, we measured the time it took participants to read different types of information presented in the postevent narrative. We used reading time as an index of attention allocation. We hypothesized that if testing affected attention to critical details, the repeated-test groups should spend more time processing misleading information, relative to single-test controls. Reading time is a long standing proxy for attention allocation (Reynolds & Anderson, 1982; Tousignant, Hall, & Loftus, 1986), with longer times reflecting more attention allocated.

Method

Participants A total of 90 undergraduate students from Tufts University participated for course credit.

Materials and procedure The experimental design was a 3 (testing: related, identical, or single) × 3 (item type: consistent, control, or misleading) mixed design. The presence/type of the initial test was manipulated between subjects, whereas item type was manipulated within subjects.

The witnessed event and test materials were the same as in Experiment 1. Narratives and test questions were constructed such that eight consistent, eight control, and eight misleading trials were included. As in Experiment 1, item type was also counterbalanced. Participants in each group first watched the video. Those in the identical and related-test groups then took an immediate cued-recall test, whereas single-test participants played Tetris. Participants were instructed to answer every question. All participants then completed the same demographic questionnaire and synonym and antonym vocabulary test as in Experiment 1. Following that phase, all participants were presented with the written narrative. The critical change in this experiment was the presentation format of the narrative. The narrative was presented sentence by sentence, to allow for the collection of reading time data. The narrative presented information that was consistent, neutral, or misleading with regard to the video. Eighty-eight sentences were presented to participants; however only 24 of those contained details of interest. Those 24 sentences were constructed to have an average length of 15 words ($SD = 3.01$) and housed the critical detail near the end of the sentence. After each sentence was presented, participants were instructed to press the space bar to move onto the next sentence. Reading time was measured as the time between each press of the space bar. After the narrative presentation, participants completed the final cued-recall test.

Results

Cued recall As with Experiment 1 we conducted a 3 (item type: consistent, control, misleading) × 3 (testing group: single, related, identical) × 2 (response: video, narrative) mixed-design analysis of variance (ANOVA) on the average proportions recalled. As Fig. 2 illustrates, responding was affected by testing, but only on misleading trials. Although the three-way interaction was not significant, $F = 1.46$, we did find a significant interaction between item type and test group, $F(4, 174) = 2.90$, $\eta_p^2 = .06$. Subsequent univariate ANOVAs collapsing across the response variable were performed in order to examine how testing affected responding for specific item types. Participants in the three testing groups did not differ in their responses given on control or consistent trials, Fs < 1; however, a significant effect of test group was found on misleading trials, $F(2, 90) = 8.34$, $\eta_p^2 = .83$. Pairwise comparisons with a Bonferroni correction demonstrated that participants in the identical-test group produced more combined video and narrative, or correct and misleading, responses than did participants in the related-test group, $t(58) = 2.31$, $d = 0.61$. In addition, participants in the related-test group produced more of these responses than did participants in the single-test group, $t(58) = 2.34$, $d = 0.63$. Thus, as Fig. 2 illustrates, testing affected the production of both correct and misleading responses, but only on misleading trials.

Reading time Table 1 presents the average reading times associated with narrative presentation in Experiment 2. What is clear from these data is that sentences that included misinformation took more time to process than did sentences

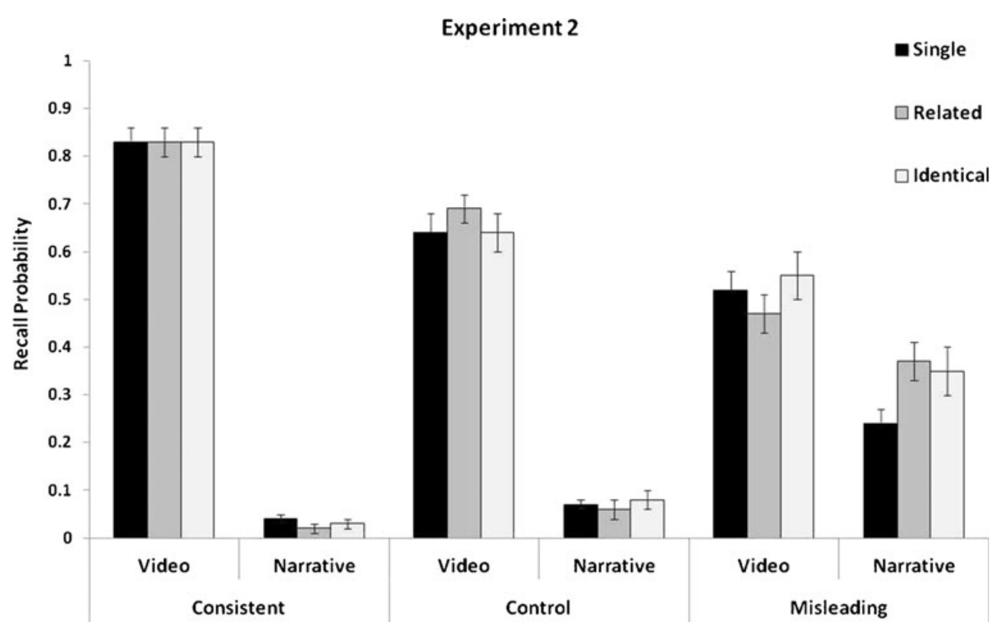

Fig. 2 Mean probabilities of responding with accurate video information or with misinformation from the narrative, as a function of test group and item type in Experiment 2

that included consistent or neutral information. This main effect was confirmed in a 3 (item type: consistent, control, misleading) × 3 (testing group: single, related, identical) mixed design ANOVA, $F(2, 174) = 26.60$, $\eta_p^2 = .23$. Pairwise comparisons employing a Bonferroni correction revealed that sentences with misinformation (M = 4,378 ms) took longer to process than did sentences with consistent information (M = 4,000 ms), $t(89) = 3.71$, $d = 0.37$. Sentences with misinformation also took longer to process than did control sentences (M = 3,690 ms), $t(89) = 6.98$, $d = 0.64$. A main effect of test group was also confirmed, $F(2, 87) = 4.77$, $\eta_p^2 = .10$. Pairwise comparisons using a Bonferroni correction revealed that slower reading times were associated with identical testing (M = 4,411 ms) than with single testing (M = 3,843 ms), $t(58) = 2.62$, $d = 0.68$. No other comparisons were significant. Finally, the interaction between test and item type was significant, $F(4, 174) = 3.48$, $\eta_p^2 = .07$.

As is demonstrated in Table 1, participants in the identical-test group took significantly longer to process sentences containing misinformation than sentences containing consistent, $t(29) = 3.16$, $d = 0.62$, and neutral information $t(29) = 5.64$, $d = 0.95$. Participants in the related-test group also took longer to read sentences with misleading details than control sentences, $t(29) = 5.01$, $d = 0.60$, and marginally longer times for misleading than consistent, $t(29) = 2.01$, $p = .04$. Finally, for single-test group, there was a small difference in reading times associated misleading and control trials, but this comparison failed to reach statistical significance. These analyses suggest that the type of immediate test may have an important effect on the processing of misleading details. Specifically, the largest difference between control and misleading sentence processing was found in the identical-test group. Three separate secondary 2 (testing: identical – related, identical – single, related – single) × 2 (item type: control, misleading) ANOVAs confirmed that the magnitude of the difference in reading times was greater in the identical-test group than in the related-test group, $F(1, 58) = 5.35$, $\eta_p^2 = .08$, and greater in the identical-test group than in the single-test group, $F(1, 58) = 10.55$, $\eta_p^2 = .15$. Finally, this interaction was only marginally significant when the related- and single-test groups were compared, $F(1, 58) = 2.21$, $\eta_p^2 = .05$.

Table 1 Mean reading times in milliseconds, by item type and group

	Item Type		
	Consistent	Control	Misleading
Experiment 2			
Single	3,850 (161)	3,678 (166)	4,003 (154)
Related	3,827 (150)	3,505 (185)	4,104 (178)
Identical	4,323 (175)	3,884 (199)	5,025 (238)
Experiment 3			
Single		4,121 (212)	4,584 (245)
Related		3,566 (125)	4,133 (146)
Identical		3,730 (148)	4,822 (214)

Standard errors are in parentheses

Discussion

Similar to Experiment 1, Experiment 2 demonstrated that immediate testing affected final-test responses on misleading trials. In Experiment 2, participants who took an immediate test were more likely to report either information from the video or the narrative than participants who did not take an initial test. In addition, the reading time results in Experiment 2 provide some support for the attention allocation hypothesis. That is, immediate testing affected the processing of details in the postevent narrative. The degree of the effect was dependent on type of test. Specifically, all groups demonstrated a difference in reading times between sentences that included misleading as compared to neutral details. The magnitude of that difference was dependent on testing condition. This difference was smallest when participants did not take an immediate test, and was greatest when the immediate test was identical to the final test and tested details that were later manipulated in the narrative. These data suggest that information that may be initially retrieved, and then later contradicted in the narrative, may capture attention and thus slow reading. These ideas will be elaborated upon in the General Discussion.

Experiment 3

Results from Experiment 2 suggest that immediate testing resulted in changes in the allocation of attention to details in a posttest narrative. These findings are novel and offer a new approach to the study of misinformation, testing effects, and RES. As such, one of the primary goals of Experiment 3 was to replicate the findings of Experiment 2. Identical-test, related-test, and standard misinformation groups were again compared. The allocation of attention within the postevent narrative was again measured using reading time data. In addition to this replication, Experiment 3 addressed the question of whether this shift in narrative processing also affected access to originally encoded memories.

Research using the RES method has demonstrated that immediate testing affected not only the later production of misinformation on a final test, but also the accessibility of the originally learned information. Changes in accessibility resulted in significant reductions in accuracy on misleading trials. Chan et al. (2009) hypothesized that immediate testing reactivates participants' memories for the original event. Reactivation places those memories into a labile state, resulting in increased susceptibility to interference and accessibility disruption. In support of this argument, Chan et al.

(2009) found that participants in an identical-testi group were less likely to retrieve the originally witnessed details on a modified–modified free recall (MMFR; Barnes & Underwood, 1959) test after misinformation than after no misinformation. In contrast, participants in a standard misinformation group were as likely to retrieval correct details after misinformation as after no misinformation under the same final-testing constraints. Thus, the testing group demonstrated disruptions in accessibility that were ameliorated by an MMFR test in the standard misinformation group.

Chan et al. (2009) argued that accessibility disruption in the misinformation paradigm is a consequence of immediate retrieval. However, accessibility disruption has not been consistently demonstrated in RES studies. For example, Thomas et al. (2010) found that when participants were warned about the validity of the narrative before the final test, accessibility disruption was diminished. Similarly, in Experiment 2 of the present study, participants did not demonstrate accessibility disruption for memory for the original event after misleading information, even though they did demonstrate greater misinformation production. That being said, accessibility disruption for memory for the original event after immediate testing was found in the present Experiment 1.

One of the goals of Experiment 3 was to further investigate original memory accessibility in a RES paradigm. In Experiment 3, we employed an MMFR final test similar to that used by Chan et al. (2009): Participants were encouraged to retrieve memories from the original video as well as memories from the postevent narrative. We hypothesized that immediate testing would direct attention resources to new information in the narrative, thereby improving encoding and strengthening the representation of that information in memory. The results from Chan et al. suggest that under the constraints of MMFR testing, participants who take an immediate test learn posttest information better, but have reduced access to the original information when compared to those who did not take an initial test. We tested this hypothesis by comparing circumstances in which participants were able to retrieve only the original information, only the narrative information, or both the narrative and video information.

Method

Participants A total of 121 undergraduate students from Tufts University participated for course credit.

Materials and procedure Consistent trials were not relevant for any hypothesis in Experiment 3, so these trials were removed. Thus, the experimental design was a 3 (testing: related, identical, single) × 2 (item type: control, misleading) mixed design. The presence/type of the initial test was manipulated between subjects, whereas item type was manipulated within subjects.

The materials and procedure of Experiment 3 were similar to those used in Experiment 2. The only difference in procedure was the final testing, and the only difference in materials was the elimination of consistent item trials. Instead, the experiment included 12 control and 12 misleading trials. Furthermore, in this experiment, Set A and Set B questions were counterbalanced, such that half of the participants received Set A questions on the initial test, whereas the other half received Set B questions on the initial test. Item type was also counterbalanced.

After video presentation followed by immediate testing or a filler task, and the distractor phase, participants were presented with the narrative. The narrative was presented sentence by sentence, and the average critical sentence length was 15 words (SD = 3.2). After the narrative, all participants completed the same demographic questionnaire and synonym/antonym test, as in Experiment 1. Finally, all participants were given the MMFR test (Barnes & Underwood, 1959). Under these testing constraints, participants were encouraged to provide both the item from the video and the item from the narrative. This procedure allowed us to examine the effect of repeated testing on original memory accessibility. Not only were we able to assess the number of misleading details reported by each group on the final test, we were also able to determine whether the misleading information was supplementing, rather than replacing, original event memory. When they received the instructions for the MMFR test, participants were informed that they could potentially remember two different pieces of information in association with a given question, but to provide information from the video first. If participants only remembered one answer, they were instructed to indicate only one answer. Participants were given a practice question to orient them to the format of the MMFR, and then they began the test.

Results

Cued recall Figure 3 illustrates the recall probabilities when participants were only able to recall one answer. This presentation is analogous to the recall probabilities presented in Experiments 1 and 2, with the primary difference being how responses were categorized. In Experiments 1 and 2, participants were asked to respond with only the originally presented information from the video. If participants responded with this information, their response was correct. If they responded with information from the narrative, they provided misinformation, and their response was scored as incorrect. In Experiment 3, participants were encouraged to respond with any information that could be associated with a question. If two answers came to mind, participants were instructed to provide both; however, the likelihood that two answers would come to mind was dependent on several factors. One factor was item type. For answers associated with control items, participants were only exposed to one possible answer, so they would only have access to one response. On misleading trials, participants were exposed to two different pieces of information; however, participants could report them only if both pieces had been encoded, and if both pieces remained accessible at retrieval. In our first analysis, we examined situations in which participants provided only one response. The data displayed in Fig. 3 are based on this analysis.

A 3 (testing: single, related, identical) × 2 (item type: control, misleading) × 2 (responding: narrative-only, video-only) mixed design ANOVA revealed a main effect of item type, $F(1, 118) = 11.35$, $\eta_p^2 = .09$. Averaging across response types and test groups, participants produced proportionally more information on control ($M = .33$) than on misleading ($M = .28$) trials. We also found a main effect of responding, $F(1, 118) = 311.11$, $\eta_p^2 = .72$. Averaging across item types and test groups, we found that participants produced proportionally more video-only ($M = .46$) responses, as compared to narrative-only responses ($M = .15$). This finding is not surprising, as on control trials participants were not exposed to new information in the narrative; thus, it was highly unlikely that narrative information would be produced on control trials. The main effect of testing was also significant, $F(2, 118) = 7.55$, $\eta_p^2 = .11$. Averaging across responding and item types, we found that proportionally more information was provided by participants in the single-test group ($M = .33$) than in the related-test ($M = .31$) or identical-test ($M = .27$) groups. However, only the difference between the single and identical groups reached significance using a pairwise comparison with a Bonferroni correction, $t(78) = 3.91$, $d = 0.85$.

More importantly, the interaction between responding and testing was significant, $F(2, 118) = 4.47$, $\eta_p^2 = .07$. As can be seen in Fig. 3, narrative-only responding was unaffected by testing. That is, on both control and misleading trials, the likelihood of producing a response from the narrative was unaffected by immediate testing. On the other hand, participants in the single-test group provided proportionally more video-only responses ($M = .52$) than did participants in the related-test ($M = .46$) or identical-test ($M = .39$) groups, again collapsing across item types. The differences between the single and identical groups and between the related and identical groups reached significance using pairwise tests with a Bonferroni correct, $t(78) = 3.61$, $d = 0.86$; $t(81) = 2.68$, $d = 0.67$. The interaction between item type and testing was also significant, $F(2, 118) = 3.27$, $\eta_p^2 = .05$. Whereas no group differences emerged for information provided on control trials, participants in the single-test group provided proportionally more information on misleading trials ($M = .34$) than did participants in the related-test ($M = .27$) and identical-test ($M = .25$) groups. The differences between single- and identical-test group responding, as well as between single- and related-test group responding, reached statistical significance, $t(78) = 3.75$, $d = 0.90$; $t(77) = 2.96$, $d = 0.80$. Finally, the interaction between responding and item type was significant, $F(1, 118) = 219.49$, $\eta_p^2 = .65$; however, this effect was driven by the expectedly low level of narrative-only responding on control trials.

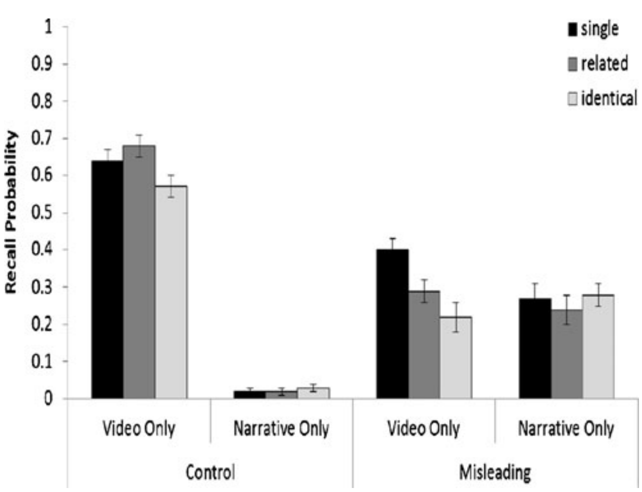

Fig. 3 Mean probabilities of responding with video-only or narrative-only information on the modified–modified free recall test, as a function of test group and item type in Experiment 3

On the surface, these data suggest that under the constraints of an MMFR test, participants in the identical-test group demonstrated greater disruptions in accessibility to information from the video on misleading trials than did participants in either of the other test groups. As Fig. 3 illustrates, video-only responding on misleading trials was significantly lower for participants who took an immediate test than for those who only took one, final test. This finding is supported by the interaction between responding and testing. However, to conclude that testing disrupted access to video information on misleading trials is premature.

Importantly, under the constraints of an MMFR test, participants are encouraged to respond with *multiple* answers. Although the analysis described above was performed to examine situations in which participants could not produce two answers, it purposely omitted situations in which participants were able to provide both pieces of information. Participants had the opportunity to provide two pieces of information only on misleading trials; therefore, the following analysis is confined only to these trials. A one-way ANOVA comparing test groups on their ability to produce both the narrative and video responses on misleading trials showed a main effect of testing, $F(2, 118) = 10.70$, $\eta_p^2 = .15$. Participants in the identical-test group ($M = .38$) produced more video and narrative responses than did participants in the related-test group ($M = .25$), $t(81) = 2.26$, $d = 0.48$. Likewise, participants in the related-test group produced more video and narrative responses than did those in the single-test group ($M = .13$), $t(77) = 2.38$, $d = 0.60$. These data suggest that immediate testing helped participants learn both the video and conflicting narrative information. As Fig. 4 illustrates, the amount of learning, as measured by "both" responding in the MMFR test, was dependent on the type of initial recall.

Reading time Reading times were examined using a 3 (testing: related, identical, or single) × 2 (item type: control, or misleading) mixed design. A main effect of item type was found, $F(1, 118) = 50.29$, $\eta_p^2 = .30$: Participants spent more time reading misleading sentences (M = 4,514 ms) than control sentences (M = 3,797 ms) (see Table 1). In addition, we found an interaction between item type and test, $F(2, 118) = 3.89$, $\eta_p^2 = .06$. Planned comparisons with a Bonferroni correction showed that participants spent more time processing sentences with misleading details than control sentences in both the identical-test group, $t(41) = 6.28$, $p < .001$, $d = 0.97$, and the related-test group, $t(40) = 4.55$, $p < .001$, $d = 0.71$. With Bonferroni correction, the difference did not reach significance for single-test participants, $t(37) = 2.16$. As in Experiment 2, we conducted three separate secondary 2 (testing: identical – related, identical – single, related – single) × 2 (item type: control, misleading) ANOVAs to examine the magnitudes of the differences in reading times between testing group pairs. This difference was greater in the identical-test group than in the related-test group, $F(1, 81) = 6.01$, $\eta_p^2 = .07$, and greater in the identical-test group than in the single-test group, $F(1, 78) = 5.32$, $\eta_p^2 = .06$. Finally, this interaction was not significant when the related- and single-test groups were compared, $F < 1$.

Discussion

Experiment 3 confirmed results in Experiment 2. Reading times were significantly slower for misleading sentences than for control sentences in the identical as compared to the other testing groups. These results suggest that testing directs attention to inconsistent information, thereby increasing the likelihood that it is better learned and remembered. Experiment 3 also confirmed that initial testing did not result in impaired access to the original event. We found that, whereas initial testing affected learning of the information in the narrative, as was shown by increased production of misleading details in the identical- and related-test groups (see Fig. 4), it did not reduce access to video information when participants were given the opportunity to produce more than one response on the final test.

Interestingly, when participants were able to access only one response, participants in the standard misinformation condition (e.g., single test) demonstrated the best memory for the original event (see Fig. 3). That is, on misleading trials, participants in this group produced the highest proportion of video-only responses. In isolation, these data would suggest that immediate testing negatively impacted access to the video information. However, a separate analysis on responses when both the video and narrative information was provided revealed that immediate testing improved access to both the video and narrative information on misleading trials. Furthermore, the amount of information retrieved was dependent on the type of initial test, with identical testing resulting in the greatest access to both video and narrative information. The implications for these results are presented in the General Discussion.

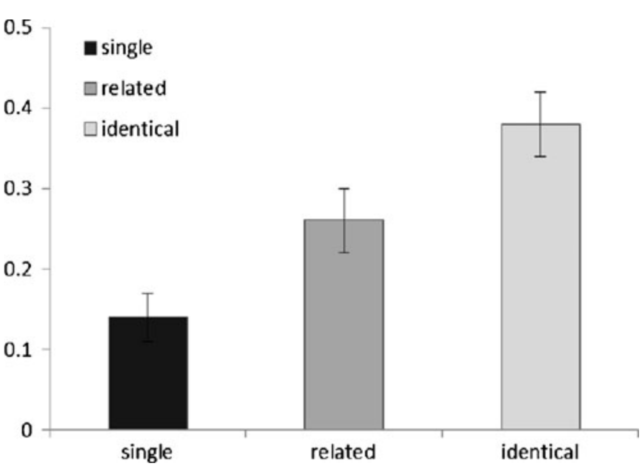

Fig. 4 Mean probabilities of responding with both video and narrative information on the modified–modified free recall test, as a function of test group on misleading trials in Experiment 3

General discussion

Research examining immediate testing in the misinformation paradigm has yielded conflicting results. When immediate testing is unconstrained, misinformation susceptibility has been shown to diminish (Gabbert et al., 2012). However, when immediate testing required participants to retrieve details later manipulated, misinformation susceptibility has been shown to increase (e.g., Chan et al., 2009). The primary motivation of the present study was to examine the complex relationship between immediate retrieval, postretrieval learning, and original memory access. By manipulating what kind of information was initially retrieved, we found that the magnitude of RES was influenced by the retrieval or attempted retrieval of details later manipulated in the narrative. Furthermore, we found that testing before misleading postevent information affected attention allocation to the postevent narrative. Finally, we found that retrieving event details prior to learning new information did not impair access to the originally learned information, but rather enhanced learning of postevent information.

Testing before misinformation

Testing of original event retention before the introduction of misinformation affected how that misinformation was processed. This finding was described as an enhanced suggestibility phenomenon, and was thought to depend on retrieval or attempted retrieval of details later distorted through misinformation. However, the present study demonstrated that retrieval or attempted retrieval of critical details was not necessary for RES. Rather, RES resulted, albeit to a lesser degree, when participants were tested on the original event, but not required to retrieve critical details. This finding is contrary to our original hypothesis. It is important to note, however, that whereas retrieval of critical details is not required for RES (as RES was observed in the related-test group), it does enhance RES. That is, larger effects were observed in the identical-test than in the related-test group. In addition, we found that under MMFR testing constraints, the participants who took an immediate test demonstrated better memory for the postevent narrative *and* the original event than did participants in a standard misinformation group, who only took one final test. These results suggest that when the final test induces a more careful search of memory, the benefits of testing may be revealed. When participants are given a forced cued-recall test as the final test, as in most RES studies, retrieval fluency may underlie responding (e.g., Thomas et al., 2010), resulting in misinformation-like effects. When final testing is less constrained, participants may be more likely to evaluate and withhold information (e.g., Gabbert et al., 2012). Research has demonstrated that more challenging retrieval tasks force participants to carefully evaluate retrieved details, resulting in improved accuracy (see Bulevich & Thomas, 2012, for a recent example).

Although the results of the present study suggest that retrieval of critical details may not be a requirement for enhanced suggestibility effects, we cannot be certain that related-test participants did not covertly retrieve these details. In fact, some researchers have contended that even in the absence of initial testing, the narrative itself serves as a form of covert retrieval (Hupbach, Gomez, Hardt, & Nadel, 2007). Moreover, according to associative memory theories such as Act-R (Anderson, 1996) and SAM (Raaijmakers & Shiffrin, 1981), materials such as those designed in the related-test group may facilitate retrieval of related information (e.g., Chan, McDermott, and Roediger, 2006). That being said, the finding that attempted retrieval of critical details was not necessary for RES fits well within a broader literature on the benefits of prior testing on later learning. For example, Wissman, Rawson, and Pyc (2011) demonstrated that the retrieval of prose material enhanced learning of subsequent related material. This study extended Tulving and Watkins's (1974) earlier work to prose material, and supports our conclusion that enhanced learning in a misinformation paradigm—termed *enhanced suggestibility*—can occur even when initial testing does not required the retrieval of later-manipulated details.

Retrieval attempts may enhance subsequent learning through a general error correction process (Carrier & Pashler, 1992; Kang et al., 2011; McClelland & Rumelhart, 1985). When participants produce a response that is later contradicted, the discrepancy may produce an error signal, activating a general error correction mechanism. This mechanism may influence encoding strategies directed to new learning (Carpenter, 2012; Pyc & Rawson, 2012). The reading time data in Experiments 2 and 3 support the conclusion that discrepancies in originally retrieved information and postevent information may influence attention and encoding. That is, because the identical-test groups spent more time reading sentences that introduced new (e.g., misleading) information subsequent to the initial test, it is plausible that a more active encoding process was engaged on these trials. However, whereas the present study suggests that changes in attention allocation may influence RES, our data also suggest that this is not the only mechanism that results in enhanced suggestibility. Similar levels of misinformation production were found in the related-test groups and in the identical-test groups, yet the reading time patterns were dissimilar. Participants in the related-test groups evidenced a reading time pattern closer to that observed in the single-test groups. In any case, it appears that attention may moderate the RES effect.

A second possible explanation for the finding that testing previously learned information assists in the learning and retention of later information is that intermittent testing prevents the buildup of proactive interference (Szpunar, McDermott, & Roediger, 2008). This proposal suggests that the benefit provided by testing after each list allows participants to segregate lists and distinguish them from one another during final recall. In other words, intermediate testing increases source discrimination (Johnson, Hashtroudi, & Lindsay, 1993). This proposal, of course, assumes that both old and new information are retained in memory. We know from the present results that this holds true in the RES paradigm.

Retrieval-enhanced learning

Perhaps it is best to reconceptualize retrieval-enhanced suggestibility as *retrieval-enhanced learning*. The present findings suggest that retrieval may not impair memory for the original event, but rather may enhance learning of new information. Evidence of this conclusion comes from the MMFR data. Participants in the repeated-testing groups were more likely to remember original and new details than were the standard misinformation participants. That being said, in the present study we did not directly examine whether participants remembered the source of details retrieved. Participants were instructed to respond with information from the video first; however, a careful source-monitoring test was not performed. It is possible that participants may have remembered two different details but were uncertain as to the source of those details. However, research has supported the conclusion that MMFR testing actually encourages source discrimination

(Hicks & Marsh, 1999; Marsh, Landau, & Hicks, 1996). Although an MMFR task may encourage source discrimination, the present study did not require source recollection. Thus, in the present MMFR test, it is possible that discrimination was not necessary.

In a recent article, Chan, Wilford, and Hughes (2012) found that initial testing did *not* enhance source memory, but rather increased the likelihood of participants erroneously reporting that misinformation was presented in both the narrative and video. These data suggest that participants may incorporate details from the narrative into their memory for the video. However, when participants were required to designate either the narrative or the video as the source of a particular memory, and the "both" response was no longer an option, participants in the repeated-test group attributed more misleading details to the narrative than did participants in the standard misinformation group. The results of Chan et al. (2012) are consistent with the present finding that when participants take a test before receiving new information, they are better able to learn that new information. The present study extends Chan et al.'s (2012) findings by demonstrating that the enhanced learning of new details from the narrative does not reduce accessibility of originally learned details. Rather, new details may be assimilated into memory for the originally witnessed event, resulting in a memorial association between a specific cue and two possibly conflicting targets.

The RES effect, as measured by a decrease in production of details associated with the original event on a final test, was greatest in the identical-test group (Exp. 1). These results are consistent with previous RES effects, and suggest that immediate identical testing may increase misinformation effects by disrupting access to the original event. RES, and misinformation effects, can also be measured by examining the proportions of misleading details that participants produce on a final test. In three experiments, we found that immediate testing resulted in greater misinformation production than what we found when participants took only one final test. That is, when participants were incorrect on misleading trials, they were more likely to answer questions with details from the narrative if they took an immediate test (Exps. 1 and 2 only). Furthermore, in Experiment 3, in which participants had the opportunity to include both video and narrative details on a final MMFR test, greater proportions of both of these kinds of details were provided by participants who took some kind of immediate test. These results support an error correction theory for enhanced posttest learning (cf. Carrier & Pashler, 1992; Kang et al., 2011; McClelland & Rumelhart, 1985).

The present results also suggest that access to originally learned information may be dependent on the type of immediate test, the way in which new learning is presented, and the options given to participants on the final test. In Experiment 1, when immediate and final test were identical, participants demonstrated the greatest misinformation effect. They were less accurate on misleading trials and more likely to produce misinformation. However, the phenomenon was mitigated in Experiment 2 when the narrative presentation was shifted from self-paced pages presenting multiple paragraphs to self-paced single sentences. These results suggest that narrative presentation may influence RES. Sentence-by-sentence presentation may have influenced the encoding of contextual and perceptual cues that were useful in segregating video and narrative information. When participants were given an MMFR test after self-paced learning (Exp. 3), we found the strongest evidence for retrieval enhanced learning. That is, immediate testing improved memory access to both the original and postevent information. This was true in both kinds of testing conditions, but strongest when the immediate and final tests were identical.

Conclusions

Does retrieval help or hurt eyewitness memory? The answer to this question is nuanced. Retrieval has both positive and negative consequences in this context. First, retrieval promotes new learning. Unfortunately, when new and incorrect information is presented, that may introduce memory reporting errors. Across three experiments, we found that participants who engaged in immediate retrieval were more likely to produce misleading details on a final test, regardless of whether the details retrieved on immediate test were later contradicted in the narrative. That being said, by changing the nature of the final test and giving participants the opportunity to report multiple targets in association with a particular cue, a different pattern emerged: Immediate testing enhanced learning of misinformation without impairing access to memory for the original event. We can conclude from the present findings that individuals who witness an event, are tested on the event, and then receive misinformation about the event in the form of a narrative spend more time processing the misinformation in the narrative than do individuals who were not given an initial test. This difference in processing results in enhanced learning.

Author note We thank Jason C. K. Chan and John B. Bulevich for comments on earlier drafts. We also thank research assistants Shelci Bowman, William Carroll, Paul Cernasov, Caroline Chen, Darius Izadpanah, and Alex Schmider for data collection and scoring.

References

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51,* 355–365.

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review, 89,* 708–729. doi:10.1037/0033-295X.89.6.708

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97–105. doi:10.1037/h0047507

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123–144). Hillsdale: Erlbaum.

Bulevich, J. B., & Thomas, A. K. (2012). Retrieval effort improves memory and metamemory in the face of misinformation. *Journal of Memory and Language, 67,* 45–58.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1563–1569. doi:10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552. doi:10.1037/a0024140

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21,* 279–283.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642. doi:10.3758/BF03202713

Chan, J. C. K., & Langley, M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 248–255.

Chan, J. C. K., & LaPaglia, J. A. (2011). The dark side of testing memory: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied, 17,* 418–432.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135,* 553–571. doi:10.1037/0096-3445.135.4.553

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science, 20,* 66–73. doi:10.1111/j.1467-9280.2008.02245.x

Chan, J. C. K., Wilford, M. M., & Hughes, K. L. (2012). Retrieval can increase or decrease suggestibility depending on how memory is tested: The importance of source complexity. *Journal of Memory and Language, 67,* 78–85. doi:10.1016/j.jml.2012.02.006

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior, 33,* 298–307.

Gabbert, F., Hope, L., Fisher, R., & Jamieson, K. (2012). Protecting against misleading post-event information with a Self-Administered Interview. *Applied Cognitive Psychology, 26,* 568–575.

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,* 1195–1209.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning and Memory, 14,* 47–53. doi:10.1101/lm.365707

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340. doi:10.1037/0096-3445.110.3.306

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

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology, 103,* 48–59.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331,* 772–775. doi:10.1126/science.1199327

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language, 67,* 17–29. doi:10.1016/j.jml.2012.02.004

Kelley, C. M., & Lindsay, D. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1–24.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639. doi:10.1037/0033-295X.100.4.609

Lane, S. M., Mather, M., Villa, D., & Morita, S. (2001). How events are reviewed matters: Effects of varied focus on eyewitness suggestibility. *Memory & Cognition, 29,* 940–947.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning and Memory, 12,* 361–366. doi:10.1101/lm.94705

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1996). The post-information effect and reductions in retroactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1296–1303.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159–188. doi:10.1037/0096-3445.114.2.159

McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition, 17,* 423–434.

Memon, A., Zaragoza, M., Clifford, B. R., & Kidd, L. (2010). Inoculation or antidote? The effects of cognitive interview timing on false memory for forcibly fabricated events. *Law and Human Behavior, 34,* 105–117.

Pyc, M. A., & Rawson, K. A. (2012). Why is test–restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 737–746. doi:10.1037/a0026166

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. doi:10.1037/0033-295X.88.2.93

Reynolds, R. E., & Anderson, R. C. (1982). Influence of questions on the allocation of attention during reading. *Journal of Educational Psychology, 74,* 623–632.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology, 84,* 171–199.

Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology: Applied, 8,* 127–142. doi:10.1037/1076-898X.8.2.127

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392–1399. doi:10.1037/a0013082

Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language, 63,* 149–157. doi:10.1016/j.jml.2010.04.004

Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading post-event information. *Memory & Cognition, 14,* 329–338. doi:10.3758/BF03202511

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior, 13,* 181–193. doi:10.1016/S0022-5371(74)80043-5

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18,* 1140–1147. doi:10.3758/s13423-011-0140-7