Journal of Memory and Language 83 (2015) 140–151

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Looking for answers in all the wrong places: How testing facilitates learning of misinformation

Leamarie T. Gordon [a,*], Ayanna K. Thomas [a], John B. Bulevich [b]

[a] Department of Psychology, Tufts University, 490 Boston Ave, Medford, MA 02155, United States
[b] Department of Psychology, Richard Stockton College of New Jersey, United States

ARTICLE INFO

Article history:
Received 27 July 2014
revision received 26 January 2015

Keywords:
Misinformation
Retrieval Enhanced Suggestibility
Testing
Test-potentiated learning
Attention

ABSTRACT

Research has consistently demonstrated that taking a test prior to receiving misleading information increases eyewitness suggestibility (Chan, Thomas, & Bulevich, 2009). Retrieval Enhanced Suggestibility (RES) is characterized by two typical findings: (1) reduced access to the originally witnessed event, which has been contextualized within a reconsolidation framework (e.g., Chan & LaPaglia, 2013), and (2) increased production of misleading post-test narrative details, which has been discussed as an example of test-potentiated learning (Gordon & Thomas, 2014). The present study focused on this latter finding, and examined attention as a potential factor moderating the relationship between initial testing and narrative detail production. We hypothesized that initial testing would influence endogenous attention allocation to details in the post-event narrative, thereby impacting narrative learning. We found that compared to participants who did not take an initial test, participants who took an initial test focused efforts on processing narrative details that were directly relevant to initial test questions, and this differential attention allocation was directly related to final test misinformation production. These experiments provide support for the view that testing enhances susceptibility to misinformation by potentiating learning of post-event information. In addition, the present findings support the conclusion that initial test questions and answers function as endogenous cues to direct attention during post-test learning.

Introduction

It is well-established that participants who are exposed to misinformation after witnessing an event are less accurate on a final test of memory, and more likely to report misleading details, than participants not exposed to misinformation (see Frenda, Nichols, & Loftus, 2011 for a review). Recent research has demonstrated that when witnesses are immediately questioned following an event, they can become even more susceptible to such misinformation. For example, Chan, Thomas, and Bulevich (2009) found that participants who took an immediate test after watching a video depicting fictitious criminal activity were less accurate on a final memory test, and more likely to report misinformation about the video, than those who did not take an immediate test. This counterintuitive pattern, termed *Retrieval Enhanced Suggestibility*, or RES (Chan et al., 2009), appears to conflict with the beneficial effects of testing that are typically observed on memory performance. In numerous situations, immediate testing enhances long-term retention of retrieved material (see Roediger & Butler, 2011 for review). Moreover, retrieval has been shown to potentiate post-test learning (Arnold

* Corresponding author.
E-mail addresses: leamarie.gordon@tufts.edu (L.T. Gordon), ayanna.thomas@tufts.edu (A.K. Thomas), John.Bulevich@stockton.edu (J.B. Bulevich).

http://dx.doi.org/10.1016/j.jml.2015.03.007
0749-596X/

& McDermott, 2013; Izawa, 1966; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Szpunar, McDermott, & Roediger, 2008; Tulving & Watkins, 1974), even in the absence of retrieval success (Richland, Kornell, & Kao, 2009). In contrast to these test-related benefits, immediate testing in the misinformation paradigm increases the misinformation effect – a seemingly negative consequence of testing. In the present study we investigated whether test-enhanced misinformation susceptibility may result from test-induced attention shifts to post-test information.

The present study examined the impact of retrieval practice on subsequent narrative learning in the RES paradigm. This approach tackles the RES phenomena from the test-potentiated learning perspective, as opposed to the complete disruption perspective inherent in reconsolidation accounts of RES. To be clear, RES is characterized by two separate findings. The first is a decrease in accuracy on a final memory test. The second is increased production of misleading details. That is, when individuals are incorrect on the final test, they report misinformation rather than some other wrong answer. Several studies suggest that these RES findings result from a reconsolidation-like mechanism (cf., Nader & Hardt, 2009) whereby access to original event information is disrupted by testing. This account suggests that immediate retrieval of the original event activates specific event details, and places them into labile states where they are highly susceptible to interference from contradictory post-event details. As the original details are no longer accessible, the misleading details are the only ones available in memory during later retrieval attempts. Although this account explains both the accuracy and production components of RES, and has received support (Chan & LaPaglia, 2013; Chan, Wilford, & Hughes, 2012; Chan et al., 2009), several studies have found that that immediate testing does not always render the original event information inaccessible (i.e., Gordon & Thomas, 2014; Thomas, Bulevich, & Chan, 2010; Thomas, Chen, Gordon, & Tenbrink, submitted for publication). For example Thomas et al. (2010) found that when participants were warned about the veracity of the narrative before final testing, those who had taken an initial test before presentation of the narrative were *more* likely to retrieve original event information on the final test than participants who had not taken an initial test. Further, Gordon and Thomas (2014) demonstrated that when participants were encouraged to report two answers in association with a given final test question, participants who had taken an initial test were more likely to report details from the narrative *and* video compared to participants who had not taken an immediate test. These results provide strong evidence against a reconsolidation/inaccessibility account.

It is also important to note that the reduced final test accuracy after testing is not consistently found within RES studies (Chan & Langley, 2011; Chan et al., 2012; Wilford, Chan, & Tuhn, 2013). Increased production of misleading details after testing, on the other hand, is *always* found. Thomas and colleagues have argued that immediate testing affects how post-event narrative information is processed, and ultimately learned. They suggest that the RES hallmark of increased production of misinformation is an indirect example of test-potentiated learning. That is, after taking an initial test, new information presented in the narrative is better learned, and thus more likely to be produced on the final test of memory. Test-potentiated learning in RES studies has been demonstrated by both increased production of misleading details as well as enhanced learning of details related to, but not directly queried by, immediate test questions (cf., Gordon & Thomas, 2014). Thomas and colleagues propose that test-potentiated learning of the narrative gives the impression of an increased misinformation effect, because at final test when original event information is required, the ease with which test-potentiated narrative learning comes to mind biases responding even if original event details are still available in memory (cf., Baddeley, 1982; Jacoby & Dallas, 1981; Jacoby, Kelley, & Dywan, 1989).

Several findings support this conclusion. First, Thomas et al. (2010) demonstrated that participants who took an immediate test demonstrated quicker decision times when choosing misleading details on a final four-alternative forced choice test compared to participants in a standard misinformation group. Second, Thomas et al. (2010) found that higher confidence judgments accompanied final test answers on misleading and consistent trials for participants who had taken an immediate test as compared to participants who had not been immediately tested. Research has consistently demonstrated that the ease with which information comes to mind serves as an indicator for confidence (Kelley & Lindsay, 1993; Nelson & Narens, 1990). Third, when participants were encouraged to discount retrieval fluency, those that had taken an immediate test demonstrated better memory for the original event details as compared to those in a standard misinformation group (Gordon & Thomas, 2014; Thomas et al., 2010).

Thus, in contrast to the reconsolidation account, Thomas and colleagues suggest that RES can be explained as a bias in responding resulting from retrieval fluency. The present study was designed to examine the mechanism by which retrieval fluency of narrative details accumulates. Gordon and Thomas (2014) suggested that immediate testing may encourage targeted shifts in attention to post-event, test-relevant details presented in a subsequent written narrative. Supporting this hypothesis, they found that participants who took an initial test spent more time reading narrative sentences that included misleading details than participants who did not take an initial test. Further, participants who took an initial test were more likely to produce misleading details on a final test of memory as compared to those who did not take an initial test. Gordon and Thomas argued that targeted shift in attention resulted in better learning of misleading details. This explanation is consistent with similar accounts found in the verbal-learning and education literature, which point to a positive relationship between attention and learning. For example, research has demonstrated that when questions targeting specific material are embedded into text passages, individuals spend more time reading the passages, and demonstrate potentiated learning of question-relevant material (Reynolds & Anderson, 1982). More recently, Szpunar, Khan, and Schacter (2013) found that interpolated tests between online lecture segments helped students sustain attention, leading to improved learning of the lecture material.

In the present study, we were motivated to better specify the processing component of this attention mechanism. The results of Gordon and Thomas (2014) suggest that immediate testing affects the attention allocated to post-test misinformation. However, the precise mechanism by which testing directs attention was not the focus of Gordon and Thomas. Thus, they only alluded to the source of this processing change, and did not determine what component of testing was responsible for the attention shift. The present manuscript builds on these preliminary findings by testing the hypothesis that post-test attentional shifts are guided by both the test question *and* the retrieved answer. Important to the RES paradigm is that when participants accurately retrieve details from the originally witnessed event during the initial test, those details will either be contradicted (with misleading details) or confirmed (with consistent details) in the post-test narrative. Even in instances when participants are incorrect on the initial test, the narrative always provides information directly relevant to one's performance on the initial test. Thus, taking a test after witnessing the original event, and before exposure to the post-event narrative, should fundamentally change how the narrative is processed. To that end, we propose that questions and answers on an immediate post-event test serve as symbolic endogenous cues. These cues facilitate strategic, voluntary, attention shifts to test-relevant details that are later presented in a narrative, similarly to how endogenous cues direct visual and spatial attention (e.g., McCormick, 1997; Stolz, 1996; Theeuwes, 1991; Yantis & Jonides, 1984).

We hypothesized that initial test questions would serve as cues during post-event narrative processing. These cues endogenously motivate attention shifts away from information that cannot answer previously tested questions, and toward information that can answer those questions. We further hypothesized that individuals' answers to immediate test questions would serve as additional endogenous attention cues guiding narrative processing. More than simply spending additional time reading sentences that include misleading details, we expected that participants would search the narrative for sentences that could answer previously tested questions, and that they would spend more time on sentences that include details that conflict with accurate memories of the original event.

Research has repeatedly demonstrated that endogenously cued attention can have measureable effects on later explicit memory (Hauer & MacLeod, 2006; MacDonald & MacLeod, 1998; Mulligan, 1998). In addition, RES production effects are diminished when participants are given initial free recall tests which are bereft of such cues (Thomas et al., submitted for publication; Wilford et al., 2013). In these free-recall RES studies, participants were not given specific test questions that could serve as cues to guide attention to specific details in the post-test narrative. Thus, another important aim the study was to directly link post-retrieval attention shifts, engendered by testing, to narrative learning in the RES paradigm. Toward that end, we predicted that the longer time spent processing test-relevant narrative details by the repeated test group would lead to better learning of these details, as measured by final test performance. Test-relevant details on consistent trials are details from the video presented accurately in the narrative (correct information), while test-relevant details on misleading trials are details presented inaccurately in the narrative (incorrect information). In the case of consistent trials, participants who took an initial test would direct attention to correct answers in the narrative, and learn them better, as evidenced by better final test performance on consistent trials compared to participants who did not take an initial test. Further, initial testing would guide new learning on misleading trials, such that participants who took an initial test would direct attention to incorrect information, resulting in increased production of misleading details (e.g., *new* information in the present context) on the final test. Importantly, production would be directly related to the amount of time participants spent processing those details.

These hypotheses were guided by the findings from Gordon and Thomas (2014), from early research on embedded questions that have shown that people pay greater attention to material when there are material-relevant questions interspersed into the material (e.g., Reynolds & Anderson, 1982), and from recent verbal-learning studies that demonstrate test-potentiated learning effects (e.g., Arnold & McDermott, 2013; Szpunar et al., 2008). Gordon and Thomas (2014) further suggested that conflicts between initial test answers and narrative information would capture attention. The present study expands this idea into a test-based attention framework, in which test questions and answers interact to flexibly guide attention toward relevant information and away from irrelevant information. It is important to note that in embedded question and verbal-learning potentiation studies, the focus is on how testing may improve retention and memory accuracy. As such, interference between two or more conflicting pieces of information is generally not within the scope of these investigations. In RES studies, on the other hand, participants learn information during the initial study phase that contradicts some details presented during a subsequent study phase. Thus, participants may also detect and/or later recollect a discrepancy between what was initially encoded, and what was presented in the context of the post-event narrative. In these instances, participants must choose *one* response for the final memory test. In the context of RES studies, the correct choice is the detail presented in the context of the original event.

It is possible that detection of a conflict between the original and post-event details may further affect attentional processes. Experiment 2 directly examined level of conflict and attention to post-event details. Specifically, initial test questions in Experiment 2 included those that targeted information presented in the original event and thus could be answered correctly, and questions that targeted information that was *not* presented in the original event. Answers to these impossible questions were presented in the narrative as supplementary misleading details. Including these questions on the first test, and their accompanying answers in the second study phase (the narrative), allowed us to examine whether testing affected processing and learning of new information in a condition where memorial conflict was reduced.

Experiment 1

The primary goal of Experiment 1 was to determine whether initial testing results in a strategic shift in attention to test-relevant details during post-event information processing. Extending Gordon and Thomas (2014), we predicted the repeated test participants (those who take an initial *and* final test) would spend more time processing sentences that contained consistent and misleading details compared to neutral sentences. Moreover, this pattern would not be observed in standard misinformation participants (those who only take a final test). To examine these hypotheses, we compared reading time differences between neutral sentences and sentences that included specific details that could answer test questions (consistent and misleading) for each testing group. In another important extension of Gordon and Thomas, we investigated how answers to test questions impact subsequent attentional processes. We hypothesized that attentional shifts in the repeated test group would be influenced by initial test accuracy. That is, participants' attention would be captured by narrative details that contradicted memory for the original event. In particular, when participants correctly recalled a detail on the initial test, they would spend significantly more time processing sentences that included details that contradicted this response (misleading sentences) compared to sentences that did not contradict their response. To test this hypothesis, we compared reading times in the repeated test group for each sentence type, based on whether initial tests responses were accurate or inaccurate. Finally, we hypothesized that learning of misleading details, as measured by production of such details on the final test, would be directly related to amount of time spent processing such details in the narrative. To test this, we looked at reading times for misleading sentences, as a function of testing group and final test answer.

Method

Ethics

All participants in both experiments provided their written informed consent to participate in the study. The study was approved by the Institutional Review Board for Social, Behavioral, and Education Research at Tufts University.

Participants

A total of 62 undergraduate students from Tufts University participated in Experiment 1 for course participation credit. Four participants failed to follow instructions, thus 58 participants (31 in the standard misinformation group; 28 in the repeated test group) were included in data analysis.

Design

The experiment employed a 3 (item type: consistent, neutral, or misleading) × 2 (testing: repeated, standard) mixed design. Presence of the initial test was manipulated between-participants, while item type was manipulated within-participants. There were 33 critical details used, eleven each in the consistent, neutral, and misleading conditions.

Materials and procedure

Participants first viewed a ~40 min pilot episode of the television program *24* (the witnessed event). After viewing the video, participants in the repeated testing group took an immediate cued recall test on details from the video (e.g., Question: What did the terrorist use to knock out the flight attendant? Answer: A hypodermic syringe). A total of 33 questions were presented for 15 s, during which participants were able to type in responses. After 15 s elapsed, the next question was automatically presented. Questions were always presented in the same order, and no feedback was provided for any question. Participants in the standard misinformation group did not take an initial test, but played Tetris (a computerized falling-rock puzzle game) for the same amount of time (10 min). All participants then completed a brief demographic questionnaire, and a synonym and antonym vocabulary test (Salthouse, 1993). The total time between the video and narrative presentation lasted approximately 20 min for both groups.

Participants then read a narrative. They were told the narrative was a synopsis of the video, and instructions did not explicitly state the narrative was truthful, or that it may contain inaccuracies. The narrative was presented to participants on a computer screen in a sentence-by-sentence format. Participants were told to simply read the narrative and to press the space bar to advance to the next sentence. Importantly, the narrative contained 33 critical sentences, a third of which included details that were misleading, a third that were consistent, and a third that were neutral (neutral sentences served as a control). Misleading narrative details replaced accurate information from the video (e.g., "The plane blew up over the Baja Desert"). Consistent narrative details presented accurate information from the video (e.g., "The plane blew up over the Mojave Desert"). Neutral details presented non-specific, but still accurate, video information (e.g., "The plane blew up over the desert"). Each critical narrative sentence served equally in each of the misleading, consistent, and neutral conditions. Sentences modified both focal and non-focal details, and ranged in length from six to 23 words. After reading the narrative, all participants immediately took a final cued recall test which was identical to the initial test. Test questions specifically targeted critical details. Participants were told to report the information presented in the video. Both initial and final cued recall tests implemented a forced responding procedure, in which participants were required to produce a response to every question. Participants were encouraged to guess if they did not know the answer. Please refer to Fig. 1 for a schematic representation of the general procedure.

Results

Cued recall

All pairwise comparisons used a Bonferroni correction unless otherwise stated. For the repeated test group, on the initial recall test 57% of participants' responses were

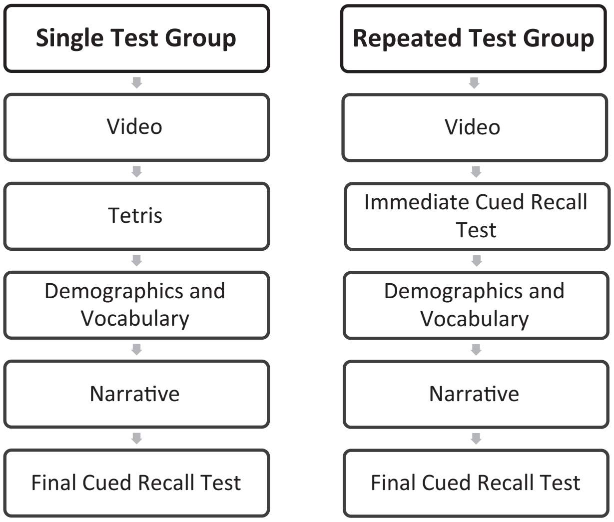

Fig. 1. General procedure for Experiments 1 and 2.

accurate and, 43% of participants' responses were incorrect. Importantly, only 3% of incorrect responses on the initial test were details that would be later presented in the narrative as misleading details.

We first report final test performance, broken down into the two components of RES: accuracy, and more relevant to the present study, production. The top half of Table 1 presents the accurate recall probabilities on the final test. Accurate recall on both Tests 1 and 2 was calculated by dividing the total number of trials in which participants produced the correct video detail out of the total number of trials for that given item type. A 3 (Item Type: Neutral, Consistent, Misleading) × 2 (Test Group: Repeated, Standard) mixed analysis of variance (ANOVA) examined the effects of item type (within subjects) and testing (between subjects) on final accurate recall. We found a main effect of item type, $F(2,112) = 57.22$, $p < .001$, $\eta_p^2 = .51$. Participants were most accurate on consistent questions ($M = .80$) as compared to neutral ($M = .60$) and misleading question trials ($M = .49$). Planned comparisons revealed that participants were significantly more accurate on consistent trials as compared to neutral trials, $t(57) = 8.43$, $d = 1.19$. In addition, participants were more accurate on neutral trials as compared to misleading trials, $t(57) = 3.37$, $d = .58$, demonstrating a standard misinformation effect. Evaluation of these significant main effects should be considered in the context of the significant item type by group interaction, $F(2,112) = 5.83$, $p < .05$, $\eta_p^2 = .10$. Participants in both groups demonstrated comparable levels of performance on neutral trials; however participants in the repeated test group demonstrated better performance on consistent trials, $t(56) = 2.65$, $d = .77$, and marginally worse performance on misleading trials, $t(56) = 1.76$, $p = .07$, than participants in the single test group.

More important to the present study, a between-participants *t*-test comparing repeated and standard test groups for misinformation production on misleading item trials found a significant difference, with repeated test participants reporting more misleading details on the final test, $t(56) = 2.74$, $d = .70$. Misinformation production was calculated by dividing the number of misleading trials in which participants produced the misleading detail out of the total number of misleading trials. Table 2 reports average misinformation production on misleading trials for each group.

Table 1
Average proportion correct on the final test in Experiments 1 and 2 as a function of item type and testing group (standard error in parentheses).

	Consistent	Neutral	Misleading-detail change
Experiment 1			
Standard	.74 (.02)	.57 (.02)	.54 (.02)
Repeated	.84 (.02)	.62 (.02)	.44 (.05)
Experiment 2			
Standard	.75 (.02)	.52 (.03)	.48 (.04)
Repeated	.93 (.02)	.72 (.03)	.48 (.05)

Overall reading time analysis. The reading time analysis was performed on a subset of the data. Only reading times for critical sentences that included neutral, consistent, or misleading details were included. Reading times faster than 300 ms were excluded from analysis. Consistent with analyses reported in Gordon and Thomas (2014) we examined average reading times in the present study. To examine differences in reading time as a function of testing group and item type, we performed a 3 (Item Type: Consistent, Neutral, Misleading) × 2 (Test Group: Repeated, Standard) mixed design ANOVA on average reading time. We interpret only the interaction between Item Type and Test Group in this analysis because we hypothesized that taking an initial test would change overall strategy during narrative reading (c. f., Britton, Piha, Davis, & Wehausen, 1978; Reynolds, Standiford, & Anderson, 1979), and hence latencies associated with all trial types, including neutral (control) trials. The interaction between item type and test group was significant, $F(2,112) = 10.31$, $p < .001$, $\eta_p^2 = .11$. Planned comparisons confirmed that while repeated test participants spent more

Table 2
Average misinformation production on the final test in Experiments 1 and 2 as a function of item type and testing group (standard error in parentheses).

	Detail change	Supplementary
Experiment 1		
Standard	.21 (.05)	–
Repeated	.35 (.07)	–
Experiment 2		
Standard	.23 (.03)	.37 (.05)
Repeated	.39 (.05)	.79 (.05)

Table 3
Average reading time for narrative sentences (in milliseconds) as a function of item type and testing group (standard error in parentheses).

	Consistent	Neutral	Misleading-detail change	Misleading-supplementary
Experiment 1				
Overall				
Standard	3428 (127)	3497 (167)	3847 (183)	–
Repeated	3289 (129)	2830 (120)	4134 (202)	–
Experiment 2				
Overall				
Standard	2917 (160)	3362 (187)	3645 (311)	3810 (242)
Repeated	3590 (223)	2833 (166)	4545 (266)	4183 (292)

time reading misleading compared to neutral sentences, $t(26) = 8.50$, $d = 1.02$, this difference was not significant in the standard misinformation group, $t(30) = 2.10$ *ns*. Similarly, participants in the repeated test group spent significantly longer time reading consistent sentences as compared to neutral sentences, $t(26) = 3.70$, $d = .67$, while there was no difference between these trial types for participants in the standard misinformation group, $t > 1$. For both groups readings times were slower on misleading compared to consistent trials [repeated test group: $t(26) = 6.06$, $d = 2.01$; standard misinformation group: $t(30) = 3.04$, $d = .49$]. Mean reading times are reported in Table 3.

Reading time contingent on Test 1. The overall pattern in reading time suggests that participants in the repeated test group spent more time processing sentences that include test-relevant information (e.g., consistent and misleading), as compared to sentences that did not (e.g., neutral). These data suggest that the test question interacts with information presented after the test. Participants strategically directed attention, as indexed by reading time, to sentences that included information that could be used to answer previously presented questions. The difference in processing time between consistent and neutral sentences, however, was not as great as that observed between misleading and neutral sentences. This suggests that although initial testing may guide attention to all test-relevant information, there is an additional factor affecting processing of misleading details. It is possible that participants are drawn to information in the narrative that not only is relevant to initial test questions, but conflicts with what they originally retrieved. To further explore this, we analyzed the reading time data from the repeated test group as a function of Test 1 accuracy. In cases where participants provided the correct video detail as the answer on Test 1, narrative sentences that included misleading details would directly contradict memory for the original event. It is these situations that we expected the largest shift in attention. This hypothesis was confirmed by the interaction found in a 2 (Item Type: Neutral, Misleading) × 2 (Test 1 Accuracy: Correct, Wrong) within-subjects ANOVA conducted on average weighted reading time,[1] $F(1,26) = 8.83$, $p < 001$, $\eta_p^2 = .25$. Planned comparisons revealed that reading times differed as a function of Test 1 accuracy on misleading trials, $t(26) = 4.79$, $d = 1.47$, but not neutral trials, $t = 1.8$ (*ns*). Participants spent more time reading misleading sentences when had reported correct answers on Test 1. A second 2 (Item Type: Neutral, Consistent) × 2 (Test 1 Accuracy: Correct, Wrong) ANOVA did not find a significant interaction, $F < 1$. Thus, reading times on consistent trials did not significantly vary as a function of Test 1 accuracy. Table 4 reports all weighted means and number of observations that influenced the weighting for each item type by group, contingent on Test 1 accuracy.

Reading time and Test 2 performance. Thus far, the data support the hypothesis that participants who have taken an immediate test differentially focus attention during post-test processing. In this final analysis using reading times, we sought to directly link differential processing to new learning, as measured by production of misleading details. In this analysis we focused only on misleading trials, as they are the only trials on which new information is presented. We performed a 2 (Test Group: Repeated, Standard) × 3 (Final Test Answer: Correct, Misleading Answer, Other Wrong Answer) mixed-design ANOVA on weighted average reading times associated with misleading trials. The interaction was significant, $F(2,82) = 4.79$, $p < .05$, $\eta_p^2 = .11$. In the repeated test group, reports of misleading details on the final test were associated with longer reading times ($M = 1434$) than were reports of other wrong, but not misleading, final test responses ($M = 796$), $t(16) = 2.50$, $p < .05$, $d = .61$. This same reading time comparison was not significant in the standard misinformation group.

Experiment 1 discussion

Experiment 1 demonstrated standard RES findings, in that compared to the standard misinformation group, participants in the repeated test group were significantly more likely to produce misleading details on a final memory test. In addition, we found that participants used Test 1 questions as a guide to learning. Repeated test participants

[1] Reading times were weighted by the number of observations in each cell. Weighting was necessary because of unequal numbers of correct and incorrect answers. For example, each participant *could* have produced eight misleading details on the final test. In the reading time analysis associated with misleading detail production, the total number of misleading details *actually* produced was divided by eight. This calculation determined the weight given to each participant's reading time mean in the analysis.

Table 4
Weighted average reading times (ms) for narrative sentences dependent on Test 1 accuracy for the repeated testing group (standard error in parentheses).

	Consistent	Neutral	Misleading-detail change
Experiment 1			
Correct			
Reading time	1658 (112)	1411 (103)	2264 (139)
Number of observations	6.07	5.55	5.55
Wrong			
Reading time	1389 (148)	1034 (137)	1088 (167)
Number of observations	4.46	3.85	3.52
Experiment 2			
Correct			
Reading time	1835 (135)	1767 (148)	2994 (259)
Number of observations	4.62	4.62	5.06
Wrong			
Reading time	2439 (261)	1297 (141)	1528 (215)
Number of observations	5	3.38	2.90

spent more time reading sentences in the narrative that could be used to answer initial test questions. This was true for both consistent and misleading narrative information, compared to neutral sentences that did not include details useful for answer test questions. Further, conflicts between original event memories, as demonstrated by correct Test 1 answers, and narrative information resulted in increased reading times. This effect was most pronounced on misleading trials. Finally, reading time analysis contingent on T2 accuracy demonstrated that misinformation production accompanied longer reading times in the repeated test group. Taken together, these findings suggest that initial test questions, answers to the questions, and discrepancies between original event memory and narrative information all impact how effectively the narrative is processed, and ultimately learned.

Experiment 2

Experiment 2 was designed to replicate and extend the findings from Experiment 1. We hypothesized that testing would influence subsequent information processing in two ways. First, test questions should guide participants to test-relevant information in the post-test narrative. We found initial support for this hypothesis in Experiment 1. Second, when narrative information directly conflicts with original event memory, participants should strategically allocate additional attention to those details. This second factor was partially support by Experiment 1. Participants in the repeated test group spent significantly more time reading misleading narrative sentences if they had produced correct answers from the video on Test 1. It is important to note, however, they did not spend more time processing consistent sentences when they had *incorrectly* produced answers on Test 1. We had not made a specific prediction regarding these trials, but it is worth noting that in this parallel situation, the information in the narrative could also have conflicted with memory for the original event. That is, when participants were incorrect on Test 1, and then encountered a detail in the narrative consistent with the video, the narrative information would not have matched what was retrieved initially. The lack of an effect of Test 1 accuracy on consistent trials suggests a boundary to the finding that initial Test 1 answers guide attention processes. That is, accuracy of answers may guide post-event attention processes only in conditions where those answers indicate well-established memories.

Test 1 answers serve as a proxy for memory strength. In Experiment 1, when participants were correct on Test 1, those responses likely signified an accessible memory representation. The indication of incorrect answers is less clear. That is, incorrect Test 1 answers could indicate failures at encoding or confusions at retrieval. In Experiment 2 we tested the hypothesis that original memory strength, as partially indexed by Test 1 answers, is a determinant of attentional processes. Specifically, the stronger the memory the more likely the resulting conflict in the narrative would engender increased attention. In order to manipulate memory strength, we presented participants with questions that could not be answered by viewing the original event. We refer to these as supplementary misleading trials. In the case of supplementary details in Experiment 2, it is not possible for an original memory to exist. Though participants were forced to provide a response, we argue that these responses do not indicate strongly represented memories of the original event. Thus, while initial test questions may lead participants to search for test-relevant information in the narrative, individuals may not spend as much time processing supplementary details because they are not truly in conflict with original memory.

We hypothesized that, as with other previously tested questions, supplementary questions would serve as endogenous cues, guiding participants to details in the narrative that could be used to answer those questions. However, we hypothesized that misleading details would receive more attention after testing than supplementary details, because on misleading trials original event memory could be directly contradicted. In contrast, on supplementary detail trials information would be provided in the narrative to answer Test 1 questions, but a conflict between the original event and the narrative could not be present. Finally, we expected that participants would report supplementary details more than misleading detail changes on the final test, and that this difference would be greater in the repeated test group compared to the standard misinformation group. That is, when participants answer questions on the final test that inquire about supplementary details, there will be no competing response (e.g., the original memory). We hypothesized that participants would disproportionately report these details as it is the only information available by which to answer final test questions. Coupled with attention shifts to test-relevant narrative information after testing, the best learning of supplementary details should be observed in the repeated test group.

Method

Participants

A total of 64 undergraduate students from Tufts University participated in Experiment 2 for course credit, and were included in analyses.

Design

The experiment employed a 2 (Test Group: Repeated, Standard) × 4 (Item Type: Consistent, Neutral, Misleading-detail change, Misleading-supplementary) mixed design. Presence of the initial test was manipulated between-participants, while item type was manipulated within-participants.

Materials and procedure

The procedure was identical to Experiment 1 (see Fig. 1); however supplementary details were included in the narrative and new questions targeting these supplementary details were introduced on both the initial and final tests. These new questions could not be answered based on the video. A new version of the narrative was constructed such that eight sentences contained misleading details that had been changed between video and the narrative, eight contained consistent information, and eight were neutral. Additionally, eight sentences presented supplementary misinformation. The inclusion of supplementary details resulted in a change to the instructions for the cued recall test. Participants were encouraged to provide an answer on all trials; however, supplementary questions on Test 1 were designed to be unanswerable. This did result in omissions. Omissions only occurred on supplementary questions associated with Test 1.

Results

Cued recall

During the initial recall test, not including performance on supplementary questions, 60% of participants' responses were accurate. Participants spontaneously produced misinformation on only 2% of Test 1 questions. The remaining 58% of responses on these trials consisted of errors of commission. For supplementary questions, participants provided an incorrect answer on 82% of the trials. Incorrect answers never corresponded to later-presented supplementary misinformation. The remaining responses were errors of omission.

As in Experiment 1, we first present final test performance in terms of accuracy and misinformation production. The bottom half of Table 1 presents the accurate recall probabilities on the final test. A 3 (Item Type: Neutral, Consistent, Misleading-Detail Change) × 2 (Test Group: Repeated, Standard) ANOVA found a main effect of item type, $F(2,124) = 45.98$, $p < .001$, $\eta_p^2 = .42$. Participants were most accurate on consistent questions ($M = .84$) as compared to neutral ($M = .62$) and misleading question trials ($M = .47$). Planned comparisons revealed that consistent trials accompanied significantly greater accuracy compared to neutral trials, $t(63) = 7.26$, $d = .99$. Neutral trials resulted in significantly better performance than misleading detail change trials, $t(63) = 3.16$, $d = 0.55$. This main effect should be considered in the context of a significant item type by test group interaction, $F(2,124) = 3.92$, $p < .05$, $\eta_p^2 = .06$. The interaction was driven by the difference between standard misinformation and repeated testing groups on consistent trials, $t(62) = 4.66$, $d = .1.19$, and neutral trials, $t(62) = 3.63$, $d = .90$.

Misinformation production was analyzed using a 2 (Production Type: Detail Change, Supplementary Detail) × 2 (Test Group: Repeated, Standard) mixed design ANOVA. We found a main effect of production type, $F(1,62) = 56.74$, $p < .001$, $\eta_p^2 = .48$. Average production was greater for supplementary details ($M = .58$) as compared to when details were changed ($M = .31$). We also found a main effect of test group, $F(1,62) = 30.13$, $p < .001$, $\eta_p^2 = .32$. Production of misleading details was greater in the repeated test group ($M = .59$) as compared to the standard misinformation group ($M = .30$). Finally, the interaction between test group and production type was significant, $F(1,62) = 13.05$, $p < .005$, $\eta_p^2 = .18$. As can be seen in Table 2, both groups produced more supplementary details than changed details; however, the difference between changed and supplementary details was significantly greater for participants in the repeated test group as compared to those in the standard test group.

Overall reading time analysis. As with Experiment 1, reading time analysis was performed on a subset of the data. Only reading times for critical sentences that included neutral, consistent, or misleading details were included. To examine differences in reading time as a function of testing group and item type, we performed a 4 (Item Type: Consistent, Neutral, Misleading-Detail Change, Misleading-Supplementary) × 2 (Test Group: Repeated, Standard) mixed design ANOVA on average reading time. As with Experiment 1, we primarily focused on the interaction. This choice was motivated by the hypothesis that initial testing would globally change processing of the post-test narrative. The interaction was significant, $F(3,186) = 6.02$, $p < .01$, $\eta_p^2 = .10$. To decompose this interaction we performed a series of mixed-design ANOVAs. All analyses were corrected for alpha inflation. A 2 (Item Type: Neutral, Misleading-Detail Change) × 2 (Test Group: Repeated, Standard) ANOVA yielded a significant interaction, $F(1,62) = 13.82$, $p < .001$, $\eta_p^2 = .18$. Participants in the repeated test group spent more time reading sentences that included misleading details compared to neutral sentences, $t(31) = 7.32$, $d = 1.36$. The difference in reading time was not significant for participants in the single test group. A 2 (Item Type: Neutral, Misleading-Supplementary) × 2 (Test Group: Repeated, Standard) ANOVA on average reading times also yielded a significant interaction, $F(1,62) = 4.78$, $p < .05$, $\eta_p^2 = .07$. Again, the difference in reading times for neutral sentences and those that included supplementary details failed to reach significance for participants in the standard misinformation group. This difference was significant for participants in the repeated test group, $t(31) = 5.01$, $d = 1.00$. A 2

(Item Type: Neutral, Consistent) × 2 (Test Group; Repeated, Standard) ANOVA on average reading times yielded a significant interaction, $F(1,62) = 14.28$, $p < .001$, $\eta_p^2 = .19$. As with previous comparisons, participants in the repeated test group spent more time reading sentences that included relevant details (in this case consistent sentences) than neutral sentences, as compared to participants in the standard group [repeated test group: [repeated test group: $t(31) = 3.78$, $d = .68$; standard test group: $t(31) = 1.81$ *ns*]. Finally, we performed a 2 (Item Type: Misleading-Detail Change, Misleading-Supplementary) × 2 (Test Group: Repeated, Standard) ANOVA on average reading times to determine if type of misinformation resulted in differences in reading time. A marginal significant effect of group was found, $F(1,62) = 3.41$, $p = .07$. No other effects were significant. Overall reading time averages for each group are reported in Table 3.

Reading time contingent on Test 1. As in Experiment 1, a 2 (Item Type: Neutral, Misleading-Detail Change) × 2 (Test 1 Accuracy: Correct, Wrong), ANOVA comparing weighted average sentence reading times found a significant interaction, $F(1,31) = 5.87$, $p < .001$, $\eta_p^2 = .16$. Planned comparisons revealed that reading times differed as a function of Test 1 accuracy on misleading trials, $t(31) = 3.75$, $d = 1.08$, but not neutral trials, $t = 1.95$ (*ns*). When participants had reported correct answers on Test 1, they spent more time reading misleading sentences. A 2 (Item Type: Neutral, Consistent) × 2 (Test 1 Accuracy: Correct, Wrong) also revealed a significant interaction, $F(1,31) = 9.23$, $p < .001$, $\eta_p^2 = .23$. Unlike in Experiment 1, a significant difference in reading time as a function of Test 1 accuracy was found on consistent trials, $t(31) = 2.71$, $d = .51$. When participants were incorrect on Test 1, they spent more time reading sentences in the narrative that were consistent with the video. All comparisons were made using a Bonferroni correction. Weighted reading time means are reported in Table 4.

Reading time and Test 2. For misleading detail change trials, we performed a 2 (Test Group: Standard, Repeated) × 3 (Final Test Answer: Correct, Misleading Answer, Other Wrong Answer) mixed-design ANOVA on weighted average reading times. As in Experiment 1, the interaction between test group and final test answer was significant, $F(2,110) = 3.74$ $p < .05$, $\eta_p^2 = .07$. In the repeated test group, reports of misleading details on the final test were associated with longer reading times ($M = 2151$) than were reports of other wrong, but not misleading, final test responses ($M = 675$), $t(31) = 5.85$, $p < .01$, $d = 1.03$. This same reading time comparison was not significant in the standard misinformation group.

For supplementary detail trials we performed a 2 (Test Group: Standard, Repeated) × 2 (Final Test Answer: Supplementary Detail, Other) mixed-design ANOVA and found a significant interaction, $F(1,51) = 20.09$, $\eta_p^2 = .28$. In the repeated test group, reports of supplementary details on the final test were associated with longer reading times of sentences containing supplementary details ($M = 3204$) than were reports of other final test responses ($M = 1112$), $t(25) = 5.08$, $p < .01$, $d = 1.00$. This same reading time comparison was not significant in the standard misinformation group.

Experiment 2 discussion

Experiment 2 replicated and extended the findings from the first experiment. As in Experiment 1, participants in the repeated test group spent more time reading sentences that contained test-relevant details compared to neutral sentences, and this pattern was not observed in the standard misinformation group. Moreover, repeated test participants' reading times were largely affected by performance on the initial test. Longer reading times were driven by the presence of information in the narrative that contradicted Test 1 responses. Finally, the results suggest that supplementary details did not capture attention to the same extent as misleading detail changes, even though participants learned supplementary details better – particularly after initial testing. These results corroborate the findings from Experiment 1, in that initial test questions, and answers to questions, impact how the post-event narrative is processed and learned. In addition, Experiment 2 revealed that the presence of a conflict between what is reported on Test 1 and what is encountered in the narrative drives processing, particularly when the Test 1 response accompanies a well-established memory.

General discussion

In two experiments, we examined how retrieval impacts subsequent information processing in the misinformation paradigm. Recently, Gordon and Thomas (2014) demonstrated that individuals pay greater attention to misinformation after testing. They argued that this differential processing enhanced learning, and increased the likelihood of accessing and reporting misleading details on later tests of memory. In the present study, several novel findings support this conclusion. First, initial test questions, and participants' responses to those questions, guided post-event narrative processing and learning. Second, the detection of a conflict between original event memory and narrative information impacted narrative reading times. This impact was tempered by strength of original event memory.

Initial test questions and answers guide processing and learning

In the context of misinformation studies, immediate post-event test questions serve as endogenous search cues, which guide processing and learning of information presented during subsequent study episodes. In the present study, after testing a greater proportion of participants' time was spent reading sentences that provided information relevant to the initial test. Less attention was directed toward uninformative information (e.g., neutral sentences). Moreover, when participants produced misleading details on the final memory test, they spent more time processing narrative sentences that contained those misleading details relative to when another response type was give on the final test. Taken together, these findings suggest that testing fundamentally changes how later information is processed and learned. Borrowing from the selective attention literature, research suggests that cognitive resources can be placed in biased anticipatory attentional states, which allow for the subsequent preferential processing of relevant information (e.g., Foxe, Simpson, & Ahlfors, 1998; Fu et al., 2001; Worden, Foxe, Wang, & Simpson, 2000). It is possible that test questions act as a form of such bias in the RES paradigm. Further evidence for the idea that testing fundamentally changes subsequent processing was provided by Pastötter et al. (2011). They had participants learn a series of word lists. Using electrophysiological measures, they found retrieval practice changed brain activity during processing of subsequent words lists, relative to a no-retrieval practice condition. Further, these changes in brain activity were directly related to later recall of the word lists. Essentially, retrieval changed subsequent encoding, which in turn predicted recall (see also Pastötter & Bäuml, 2014).

In the repeated test group, initial test performance also guided processing of the narrative. We argued that this resulted, at least in part, from the detection of a conflict between what was originally encoded and what is later encountered. Participants likely had some expectation of the post-test narrative content based on their original event memory, and memory of Test 1 responses. Thus, what they experienced potentially conflicted with what they expected, and what they truly remembered, resulting in what Whittlesea and colleagues have referred to as disfluent processing (cf. Kronlund & Whittlesea, 2006). In both Experiments when participants were correct on Test 1, indicating a strong original memory representation, they spent the most time reading misleading detail-change sentences which conflicted with these responses.

Inconsistent with this explanation, in Experiment 2 we found that when participants were *incorrect* on Test 1, they spent more time reading sentences that included details consistent with the original event compared to other sentence types. In these cases, we would have expected that reading times would *not* be longer, because incorrect Test 1 responses indicate a lack of a strong original event memory, and thus a conflict between original memory and narrative information would be unlikely. However, incorrect test answers are only indirect assessments of memory. An incorrect answer could indicate inefficient encoding, temporary inaccessibility, or even source confusions. Thus, the findings of longer reading times on consistent trials where participants were incorrect on Test 1 are uninformative, because we have no way to ascertain why participants were incorrect. It is possible that on these trials participants confused encoded elements of the original event, correctly accessing a memory representation of a given item or detail, but place that detail in the incorrect context.

Misleading-supplementary trials provide important insight here to both our hypothesis, and the across-experiment difference. First, in Experiment 2 we *knew* why participants were incorrect on Test 1 supplementary trials. Participants did not encode supplementary details while watching the video, thus a memorial conflict between original encoding and narrative information was not possible. This was evidenced by the disproportionate reporting of supplementary details relative to misleading detail changes on Test 2, because they were the *only* details participants had available in memory answer final test questions. We found that, as with all initial test questions, supplementary questions directed attention to narrative details that could answer the questions. Importantly, however, the difference in reading time between sentences that included misleading detail changes and neutral sentences was significantly greater than that observed between supplementary details and neutral details. This strongly supports our hypothesis that increased reading times result from true memorial conflict.

Second, we argue that the presence of supplementary trials in Experiment 2 drove the reading time differences on incorrect Test 1 trials observed in Experiment 2 that were not present in Experiment 1. While admittedly post hoc, we argue that the inclusion of supplementary questions may have globally affected narrative processing, resulting in across-experiment differences in both reading time *and* accuracy. Regarding reading time, all reading times in Experiment 2 were longer than those observed in Experiment 1. This increase could have been enough to reveal a significant difference in reading times as a function of Test 1 accuracy on consistent trials. In regard to final test accuracy, initial testing was associated with better final test performance in Experiment 2 as compared to Experiment 1. That is, on consistent and neutral trials in Experiment 2, repeated test participants performed about 20% better than standard misinformation participants. These differences were smaller on consistent trials and non-existent on neutral trials in Experiment 1. We argue that supplementary questions on the initial test in Experiment 2 resulted in more effortful retrieval attempts, because they were unanswerable. Borrowing from the *elaborative retrieval hypothesis* (Carpenter, 2009; Carpenter, 2011; Carpenter & DeLosh, 2006), when retrieval cues are presented to individuals, for example during the initial test phase, the cues activate information in memory related to that cue during the search for the target. When target information is not readily available, as is always the case for supplementary details, a more extensive and elaborate search of memory is necessary. This more elaborate search would likely activate more memory traces from the video, accounting for the increase in overall final test accuracy from Experiments 1 to 2. Essentially, the inclusion of difficult supplementary questions made the initial test *more effective*. Toward that end, the test questions in Experiment 2 may have been more effective in cuing subsequent attention during narrative processing.

Testing and attention

The results of the present study point to the importance of attention in test-enhanced learning of misinformation, or RES. It is important to note, however, some prior studies that have argued against attention as a key factor in test-enhanced learning. For example, Shapiro and Gordon

(2012) compared a condition in which students were given in-class questions targeting lecture material to a condition in which targeted lecture material was presented in a red flashing font and the students were explicitly told that the material was important to an upcoming exam. They found that students' exam performance was better for material that was initially tested relative to material presented in a way designed to grab attention. Richland et al. (2009) distinguished between the effects of testing and attention in a similar manner. They gave participants pretest questions and then presented a passage for study that could be used to answer those questions. Participants who were given pre-test questions demonstrated better learning on a final test compared to participants who were given the passage with important details in highlighting type-face. These studies suggest that testing may not be affecting attention to post-test information. However, in these studies, testing groups were compared to groups in which attention was manipulated exogenously (e.g., changing type-face) and not directly measured. In the present study, we compared a testing group to a non-testing group, and directly measured attention allocation to post-test material in each group. The present study directly examined the effect of testing on post-test attention allocation, whereas previous studies have only indirectly examined the relationship between testing and attention. Further, endogenous cuing via testing activates top-down processes and involves conscious control, whereas exogenous cuing such as font changes activates bottom up processes, and merely grabs attention involuntarily. Thus, it may be inappropriate to discount the role of attention in test-enhanced learning based on comparisons made between two vastly different attention direction mechanisms.

Conclusions

Although retrieval is one of the most broadly studied areas in cognitive psychology, little is known about the underlying causes of retrieval-driven potentiation effects (see Pastötter & Bäuml, 2014 for a recent review on the topic). In the case of RES, the answer may be particularly difficult to disentangle. The RES paradigm offers a unique learning situation in which post-retrieval study opportunities expose participants to incorrect information. Later, participants are expected to disregard what they have learned in order to provide correct details on a final test of memory. What is known, from the present study and prior RES studies (e.g., Chan et al., 2009; Gordon & Thomas, 2014; Thomas et al., 2010), is that test-potentiated learning *does* occur in this context. After testing, participants learn misinformation better, and have difficulty discounting the incorrect information on final memory tests. In the present study, we did not have a *direct* measure of post-test learning, because participants were only able to provide one response on the final test. That is, there lies the possibility that individuals learned the information well, but still did not report it on the final test. Nevertheless, the present set of experiments provides evidence as to what drives post-retrieval processing, and offers some insight into what drives test-potentiated learning, at least in the unique context of the RES paradigm. We identified two factors likely at play. First, initial test questions serve as endogenous cues by directing participants study efforts during post-retrieval learning. Second, exposure to information in the narrative that conflicts with a strong original memory additionally guides attention and processing resources. In the supplementary details condition, testing guided attention; however there was no original memory to compete with the relevant detail. The absence of a conflict allowed for faster processing relative to misleading detail changes, and under the same reasoning, the best condition for learning. These findings serve as an important starting point to understanding the larger picture of test-potentiated learning effects in the misinformation paradigm.

We do not contend that attention is the sole mechanism by which testing potentiates learning in the misinformation paradigm. However, we argue that initial testing guides new learning, emphasizes the presence of memory discrepancies, directs and then modulates attention. The result is enhanced learning and accessibility of post-event details, which manifests as RES. Not only are these findings theoretically important, but they have practical significance. Questions posed to a witness at the scene of a crime will influence how that witness processes new information related to that original event. Perhaps when the witness reads about the crime in a news source he will unknowingly spend more time processing details that are inconsistent with answers originally provided. Questions posed to a student after exposure to new material will similarly influence how that student processes related concepts presented after testing. The result of testing is attention to contradictory information. The result of increased attention is enhanced learning.

Acknowledgments

We would like to thank research assistants Reema Al-Marzoog, Shelci Bowman, William Carroll, Paul Cernasov, Shannon Robinson, and Danielle Sorcher for data collection and scoring.

References

Arnold, K. M., & McDermott, K. B. (2013). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review, 20*, 507–513.

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review, 89*, 708–729.

Britton, B. K., Piha, A., Davis, J., & Wehausen, E. (1978). Reading and cognitive capacity usage: Adjunct question effects. *Memory & Cognition, 6*, 266–273.

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.

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.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*, 268–276.

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. (2013). Impairing existing declarative memory in humans by disrupting reconsolidation. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 9309–9313.

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.

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.

Foxe, J. J., Simpson, G. V., & Ahlfors, S. P. (1998). Cued shifts of intermodal attention: Parieto-occipital ~10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport, 9*, 3929–3933.

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science, 20*, 20–23.

Fu, K. G., Foxe, J. J., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2001). Attention-dependent suppression of distracter visual input can be cross-modally cued as indexed by anticipatory parieto-occipital alpha-band oscillations. *Cognitive Brain Research, 12*, 145–152.

Gordon, L. T., & Thomas, A. K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory & Cognition, 42*, 186–197.

Hauer, B. J. A., & MacLeod, C. M. (2006). Endogenous versus exogenous attentional cuing effects on memory. *Acta Psychologica, 122*, 305–320.

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, 18*, 879–919.

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

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). *Memory attributions*. Hillsdale, NJ, England: Lawrence Erlbaum Associates Inc.

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.

Kronlund, A., & Whittlesea, B. W. A. (2006). Remembering after a perception of discrepancy: Out with the old, in with the two. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5*, 1174–1184.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98*, 291–310.

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 168–180.

Mulligan, N. W. (1998). The role of attention during encoding in implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 27–47.

Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews Neuroscience, 10*, 224–234.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.). *The psychology and learning and motivation* (Vol. 26, pp. 125–173). New York: Academic Press.

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology, 5*, 286.

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 287–297.

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

Reynolds, R. E., Standiford, S. N., & Anderson, R. C. (1979). Distribution of reading time when questions are asked about a restricted category of text information. *Journal of Educational Psychology, 71*, 183–190.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*, 243–257.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20–27.

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

Shapiro, A. M., & Gordon, L. T. (2012). A controlled study of clicker-assisted memory enhancement in college classrooms. *Applied Cognitive Psychology, 26*, 635–643.

Stolz, J. A. (1996). Exogenous orienting does not reflect an encapsulated set of processes. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 187–201.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 6313–6317.

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.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49*, 83–90.

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.

Thomas, A. K., Chen, C., Gordon, L. T., & Tenbrink, T. (2015). Choose your words wisely: What verbal hesitation indicates about eyewitness accuracy (submitted for publication).

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.

Wilford, M. M., Chan, J. C. K., & Tuhn, S. J. (2013). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. *Journal of Experimental Psychology: Applied*.

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band EEG increases over occipital cortex. *Journal of Neuroscience, 20*, 1–6.

Yantis, S., & Jonides, J. (1984). Abrupt onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 601–620.