Journal of Memory and Language 95 (2017) 190–199

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The forward effects of testing on eyewitness memory: The tension between suggestibility and learning

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

[a] Assumption College, United States
[b] Tufts University, United States

ARTICLE INFO

Article history:
Received 8 November 2016
revision received 25 February 2017
Available online 3 May 2017

Keywords:
Retrieval enhanced suggestibility
Interim testing
Misinformation effects
Encoding
Test enhanced learning

ABSTRACT

Research has consistently demonstrated that taking a test prior to receiving misleading information can increase misinformation susceptibility (Chan, Thomas, & Bulevich, 2009). However, research has also demonstrated that testing enhances subsequent learning (e.g., Tulving & Watkins, 1974; Wissman, Rawson, & Pyc, 2011). The goal of the present study was to examine these seemingly contradictory effects of testing. In two experiments we tested the hypothesis that testing influences how post-test information is processed. Depending on the nature of the later memory test, test-related processing can result in either memory errors or enhanced learning effects. Experiment 1 indicated that testing may result in elaborative processing of post-test material, resulting an increase in misinformation suggestibility. Experiment 2 suggested that increased suggestibility after testing may be understood as test-related learning of post-test material. Taken together, the results suggest that interim testing occurring between an original event and post-event misinformation may enhance memory suggestibility, because testing results in elaborative processing of subsequent material. However, interim testing also helps segregate memory for each source, resulting in test-potentiated learning within the misinformation paradigm.

Introduction

Eyewitness memory is often studied in the context of the misinformation paradigm, where participants witness an original event, encounter misleading post-event information, and then are tested on their memory for the original event (Frenda, Nichols, & Loftus, 2011 for review). The typical finding is that misleading post-event information negatively impacts memory for the original event. However, the standard misinformation paradigm omits an important factor that may very well influence the ability of eyewitnesses to accurately remember witnessed events. Eyewitnesses commonly engage in an initial retelling of the witnessed event to an emergency operator or on-the-scene officer. This initial retelling can be viewed as a memory test of the originally witnessed event. Considering the large literature on retrieval enhanced learning effects in verbal learning and education (for review see Roediger & Butler, 2011), a test on the witnessed event should improve retention of the witnessed event. However, research within the misinformation paradigm has consistently demonstrated that testing participants' memory immediately after they witness an event renders them more susceptible to misleading post-event information, and ultimately less accurate on subsequent memory tests. This test-related memory impairment is known as *Retrieval Enhanced Suggestibility* (RES) (Chan, Thomas, & Bulevich, 2009; Gordon & Thomas, 2014; Gordon, Thomas, & Bulevich, 2015; Thomas, Bulevich, & Chan, 2010).

In the RES paradigm, participants view a video of an event, then take an immediate memory test on the event. Next, participants read or listen to a post-event narrative. The narrative is a synopsis of the event that includes details both consistent and inconsistent with the original event. A final memory test for the original event follows. RES is typically demonstrated on this final memory test in two ways: retrieval-enhanced errors of omission and retrieval-enhanced errors of commission. When participants make errors of omission, they are less likely to recall original event details on a final memory test after exposure to misinformation in the narrative. Participants who take an interim test between the original event and the post-event narrative perform even worse compared to those who do not (e.g., Chan et al., 2009; Gordon & Thomas, 2014). This suggests that testing prior to the narrative may reduce accessibility of original event details (cf., Chan & LaPaglia, 2013). However, we caution that these retrieval-enhanced errors of omission in RES studies are not always found (see Chan & Langley, 2011; Chan, Wilford, & Hughes, 2012; Wilford, Chan, & Tuhn, 2013 for examples). Errors of commission also increase after taking

* Corresponding author at: 500 Salisbury St., Worcester, MA 01609, United States.
E-mail address: lt.gordon@assumption.edu (L.T. Gordon).

http://dx.doi.org/10.1016/j.jml.2017.04.004
0749-596X/

an interim test in the RES paradigm (Chan et al., 2009; Gordon et al., 2015; Thomas et al., 2010). When participants make errors of commission, they are not simply unable to recall details from the original event, they instead report misleading event details suggested in the narrative. As this pattern of results increases when interim testing is implemented, it is plausible that taking an interim test may impact how well narrative information is learned (cf., Arnold & McDermott, 2013; Gordon & Thomas, 2014; Wahlheim, 2015; Wissman, Rawson, & Pyc, 2011). Further, in contrast to increased errors of omission, increased errors of commission are always demonstrated after interim testing in the RES paradigm.

The present study focused on how and why taking a test on a witnessed event consistently increases errors of commission. Studying errors of commission in this context has both applied and theoretical importance. Understanding the situational factors that may increase the likelihood of an eyewitness attending to and later reporting inaccurate post-event details can inform the procedures used to interview witnesses both at the scene, and at later points during an investigation. In addition, if prior testing impacts how well subsequently presented misinformation is learned in this paradigm, then this work provides an important extension of test-potentiated learning effects to the novel context of eyewitness memory. Further, it presents a new paradigm in which to test theories of the mechanisms underlying these effects.

Forward effects of testing

Research has consistently demonstrated that interim testing, or testing between learning episodes, influences learning of post-test material. This phenomenon has been characterized as a *forward effect* of testing (Pastötter & Bäuml, 2014), and both encoding and retrieval accounts attempt to explain this effect. The present manuscript examines both encoding and retrieval accounts of RES. Before doing so, we first present prior research examining forward effects of testing.

Testing prior to restudy of a given item has been shown to facilitate performance on a subsequent test of that item (cf., Izawa, 1971; Karpicke, 2009) and also facilitate learning of new material (Wissman et al., 2011). This latter finding, the forward effect of testing, has routinely been captured in the A-B, A-D paired associate learning paradigm. In this paradigm participants study two lists of cue-target word pairs. Each list presents word pairs with the same cue word (word A), but the target word changes between List 1 (word B) and List 2 (word D). Using this paradigm, Tulving and Watkins (1974) found that when participants were tested on List 1 pairs before studying List 2 pairs, later recall of List 2 improved on both a direct test of List 2 and a modified modified free recall test (MMFR) where both List 1 and List 2 were recalled. In a misinformation effect study, which conceptually resembles an A-B, A-D paradigm, Gordon and Thomas (2014) demonstrated that including an immediate test of the originally witnessed event led to better recall of details from the post-event narrative on an MMFR test, compared to a condition where an immediate test of the original event was not included. Other recent research has replicated the forward effect of testing in a number of verbal learning and educationally relevant contexts (Arnold & McDermott, 2013; Aslan & Bäuml, 2016; Szpunar, McDermott, & Roediger, 2008; Wissman et al., 2011).

Encoding explanations for forward effects of testing

One theory proposes that testing facilitates learning of new material in verbal learning and education studies, because it improves encoding of the material. For example, encoding may be facilitated via the unconscious activation of related information during initial testing (c.f., Carpenter, 2011; Chan, McDermott, & Roediger, 2006; Grimaldi & Karpicke, 2012; Hays, Kornell, & Bjork, 2013). That is, interim memory retrieval may activate the target and target-related information. That activation in turn may facilitate the incorporation of new information into memory. Relatedly, testing may change participants' conscious encoding strategies (e.g., Wissman et al., 2011), leading participants to prioritize rehearsing or reviewing information that is related to previous test questions.

Several studies point toward encoding explanations of forward testing effects. For example, interim testing is particularly effective when post-test material is related to the tested material (cf., Gordon & Thomas, 2014). Wissman et al. (2011) demonstrated that interim testing facilitated learning of prose material that was related to previously tested material. Kornell and colleagues (Hays et al., 2013; Kornell, Hays, & Bjork, 2009; Richland, Kornell, & Kao, 2009) demonstrated that tests can facilitate subsequent study episodes of relevant information, even when initial retrieval has failed. Additional studies have linked interim testing with changes in post-test encoding strategies. An early study demonstrated that individuals spent more time reading passages after interim testing (Reynolds & Anderson, 1982). A more recent line of research has found that interim testing results in sustained attention during subsequent study and reduces mind-wandering (Szpunar, Khan, & Schacter, 2013).

The aforementioned research suggests that test-related changes in the encoding and learning of post-test material, or forward effects of testing, may also occur in the context of the misinformation effect paradigm. However, when interim testing occurs between an original event and the presentation of post-event misinformation in an eyewitness memory paradigm, participants are typically tested on memory for the original event, and post-test learning is not directly queried as in a verbal learning study. In a misinformation paradigm, interim testing leads to an increase in misinformation suggestibility, as indicated by intrusions of narrative details, or errors of commission, on the test of the original event. Thus, in the RES eyewitness paradigm, test-potentiated learning is only indirectly measured via errors of commission.

While learning of post-test information has not directly been measured in a RES study, Gordon, Thomas, and colleagues have begun to examine an encoding explanation of test-related increases in suggestibility. For example, Gordon and Thomas (2014) found that interim testing affected the amount of time participants spent reading individual sentences in the post-event narrative. The difference in reading time associated with sentences that included misleading details as compared to neutral sentences that offered no specific details was greater for participants who took the interim test compared to standard misinformation participants. Gordon et al. (2015) extended these findings by demonstrating that participants who took the interim test spent more time reading sentences that included details directly relevant to the interim test questions (either consistent with *or* contradictory to the encoding event) compared to neutral sentences. A contingency analysis based on performance on interim test questions revealed that when participants were *correct* on interim test questions, they spent more time reading details that contradicted their responses. Finally, Gordon et al. yoked narrative sentence processing times to final test output. When participants who took an interim test produced misleading details on the final cued recall test, they had spent more time processing the misleading narrative sentences that introduced those details as compared to trials where they reported some other wrong answer on the final test. This difference was not present in the standard misinformation group who did not take an interim test. Taken together, these results suggest that the inclusion of interim testing changes the encoding strategy used to process the post-event narrative.

Thus far, support for the hypothesis that interim testing affects how the post-event narrative is processed has been accumulated only via reading time data (e.g., Gordon & Thomas, 2014; Gordon et al., 2015). However, in many RES studies, (e.g., Chan & Langley, 2011; Chan et al., 2009; Thomas et al., 2010), participants listened to, as opposed to read, the post-event narrative. In these experiments test-related changes in processing time cannot be directly measured. When the narrative is presented in a written format and participants' reading times of individual sentences are measured, it is at the discretion of individual participants to decide how much time to spend processing each sentence. Alternatively, when the narrative is presented in an experimenter-paced aural manner, participants cannot choose to spend more time listening to, or processing, some sentences relative to others, without cost to other presented material. Yet, RES errors of commission still occur. These errors may remain because participants continue to process critical details even in the context of aural presentation.

In the present study our first goal was to examine whether test-related elaborative encoding may be disrupted, thereby reducing test-potentiated accessibility of narrative details. We propose that even when narratives are experimenter-paced, participants who take an interim test are cued to spend additional time thinking about, or elaborating upon, misleading narrative details, and this elaboration results in RES errors of commission. In order to test this hypothesis, we introduced a manipulation designed to impact test-related elaborative processing of narrative details. In order to manipulate participants' ability to rehearse narrative content, we capitalized on the constraints of divided attention and employed a secondary distractor task during the presentation of the narrative. This task allowed encoding of narrative details, but minimized the cognitive resources available to effectively rehearse those details. We chose a particularly demanding secondary task that shared several processing dimensions with the concurrent task in order to have a greater impact on hypothesized test-related elaborative encoding (Wickens, 2008). We hypothesized that the secondary task would reduce the likelihood of test-related elaborative encoding of critical narrative details. This reduction in test-related elaborative encoding should in turn decrease RES errors of commission.

Retrieval explanations for forward effects of testing

Retrieval explanations for forward effects of testing typically assume that, as opposed to encouraging elaborative encoding, interim testing may promote contextual list segregation of learning episodes. This, in turn, reduces interference between the pre- and post-test information (e.g., Bäuml & Kliegl, 2013; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Sahakyan & Hendricks, 2012; Szpunar et al., 2008; Weinstein, Gilmore, Szpunar, & McDermott, 2014). More information can be recalled from learning episodes that follow interim testing because individuals can better engage in retrieval monitoring of source information. In the context of the RES paradigm, the mechanism by which testing may promote learning of post-event misinformation remains debated. Although Gordon, Thomas, and colleagues have provided evidence that interim testing may influence the processing of post-event information, retrieval-based explanations for these test-potentiated effects remain a possibility.

Recent work by Jacoby and colleagues (e.g., Jacoby, Wahlheim, & Kelley, 2015; Putnam, Wahlheim, & Jacoby, 2014; Wahlheim & Jacoby, 2013) suggests that when individuals notice and later remember change between study episodes, interference between study episodes can be reduced. In a recent verbal learning study conceptually analogous to the RES paradigm, Wahlheim (2015) had participants study A-B cue-target pairs followed either by a test of A-B, or restudy of A-B. After test or restudy, participants were presented with List 2, which included A-B and new A-D pairs (same cue, different target). Following List 2, participants were given a memory test in which they were presented a cue and required to retrieve the target from List 2. After attempting to recall the List 2 target, participants indicated how certain they were that the target had had earlier changed from List 1 to List 2. Wahlheim found that change was recollected more often for tested as compared to restudied A-B, A-D pairs. Recollection of change may have provided access to representations that preserve temporal relationships of information from separate occasions, facilitating source segregation. It is possible that in the RES paradigm, testing facilitates encoding and learning of the post-event narrative because participants are more likely to notice and remember change between study episodes. Thus, the second goal of this study was to examine the extent to which interim testing promoted segregation between the original event and post-event narrative by measuring both learning of the post-event narrative, as well as memory for change.

The present study

In two experiments we investigated how forward effects of testing manifest in the context of the misinformation paradigm. In Experiment 1, we examined whether disrupting test-related elaborative encoding of narrative details would reduce RES. We also explored whether interim testing would impact participants' recollection of change in a misinformation experiment, where the goal is to assess memory for the original event.

In Experiment 2, we investigated whether interim testing truly potentiates learning of new misleading narrative details. To do this, we adopted the novel approach of examining memory for the narrative directly. In all previous RES studies (e.g., Gordon & Thomas, 2014; Gordon et al., 2015), *learning* of new narrative details was inferred by examining errors of commission. The present Experiment 2 required participants to report what they remembered from the post-event narrative only. This is akin to a test of "List 2" in a verbal learning A-B, A-D paradigm. Surprisingly, such an investigation has not yet been published, yet can be highly informative as to the mechanism by which interim testing may impact learning and monitoring of complex related event memories. In addition, testing of the narrative in Experiment 2 allowed us to compare the test-potentiated elaborative encoding account of forward effects against the retrieval-based segmentation account. The test-potentiated elaborative encoding account was examined by assessing memory for the post-event narrative under conditions where test-potentiated elaborative encoding was minimized by a secondary task. Retrieval-based segmentation was assessed by measuring change recollection after the final test under the same encoding constraints.

Experiment 1

Experiment 1 examined the relationship between RES misleading errors of commission and enhanced elaboration of the post-event narrative. We hypothesized that if increased errors of commission demonstrated after interim testing are influenced by elaboration during encoding, then reducing the opportunity for such elaboration should also reduce misleading errors of commission. To test this hypothesis, we employed a secondary task during narrative processing specifically design to reduce the likelihood of elaborative encoding. Participants' recollection of change between video and narrative content were also measured.

Method

Design

We employed a 2 (Attention: Full, Divided) × 2 (Testing: Standard, Interim) × 3 (Item Type: Consistent, Neutral, Misleading) mixed factorial design. Attention and Testing were manipulated between-subjects, while Item Type was manipulated within-subjects.

Participants

A sample size estimation was calculated using G*Power Version 3 software. Using moderate parameters (power = 0.8, effect size $f = 0.25$) the analysis estimated a sample size of 124. Sixty-one male and 99 female undergraduates from Tufts University, with a mean age of 19.3 ($SD = 2.5$), participated and were compensated either with course credit or $15. Of the total 160 participants, 59% identified as White, 15% as Asian, 6% as Black, 6% as Hispanic or Latino, 7% as Multi-racial, and 7% as another race not listed. All participants spoke English as their primary language, or indicated proficiency in English. All participants provided written informed consent prior to beginning the experiment. Participants were randomly assigned to one of four groups, with 40 participants in each group.

Materials

Witnessed event. A 22 min excerpt from the black and white silent film "Rififi" (Bezard, Bérard, Cabaud, & Dassin, 1955) was used as the witnessed event. The clip portrayed a group of four men committing a burglary in the middle of the night. No participant reported seeing this movie before.

Memory tests. Twenty-four questions querying specific details from the video were constructed as memory test stimuli. A pilot experiment in which 44 participants watched the video and then took an immediate test was used to select test questions. Only questions in which mean correct responding ranged from 0.40 to 0.70 were selected. Identical cued recall tests were used as the interim and post-narrative memory tests in the Interim Test group. Following the general procedure used in related studies in the lab, confidence ratings associated with responses (made on a scale of zero to 100) were collected during each test. Confidence ratings were conducted as part of another project and therefore, are not reported in this manuscript. A discrepancy recollection test followed the final post-narrative cued recall test. In this task participants were re-presented with each final test question. They were instructed to respond YES if they remembered different details from the video and narrative in association with each question, and NO if they did not remember different details.

Post-event narrative. The audio narrative contained 24 critical sentences that introduced consistent, neutral, and misleading information about the video (8 details each), in addition to 91 filler sentences that presented information that was not tested in either the pre or post-narrative test phases. At least three filler sentences were present between critical sentences. Consistent sentences contained details that were accurate regarding the witnessed event (e.g., *From a drawer that holds valuables, he removes a ring.*) Neutral sentences included details presented in the video, but not manipulated in the narrative (e.g., *From a drawer that holds valuables, he removes a piece of jewelry.*) Misleading sentences included details from the video that had been changed in the narrative (e.g., *From a drawer that holds valuables, he removes a necklace.*) Sentences serving as misleading, neutral, and consistent were counterbalanced across participants. The critical detail (e.g., ring/jewelry/necklace) was always present at the end of the sentence. A female voice read the narrative at a standard normal reading pace.

Procedure

After informed consent, participants watched the video. They were instructed to watch the video carefully, and informed that their memory for information presented would be later assessed. Following the video, participants in the **interim test group** took an immediate cued recall test on details from the video. Instead of taking this initial test, the **standard misinformation group** completed a Sudoku puzzle for an equivalent amount of time (6 min). All participants then completed a brief demographic questionnaire, and a synonym and antonym vocabulary test (Salthouse, 1993). Participants then listened to the audio narrative describing the events from the video. Half of the participants from both the interim test and standard misinformation groups listened to the narrative while their attention was divided by a secondary task, while the other half listened to the narrative under a full attention condition. Under full attention constraints, participants simply listened to the audio narrative. Under divided attention constraints, participants simultaneously listened to the narrative while completing a same-modality distractor task (adapted from Foerde, Knowlton, & Poldrack, 2006). Our goal was to disrupt post-encoding elaboration of narrative details that we hypothesized was increased after testing. Thus, the secondary task was designed to not interfere with critical detail encoding; however participants' ability to spend any additional time elaborating on, thinking about, or reviewing, the critical details was minimized. Importantly, after the narrator read each critical sentence, a high pitched tone signaled an upcoming series of eight high and low pitched tones. Participants counted the number of high pitched tones in the series, and recorded the number on a sheet. The number of high pitched tones following each signal randomly varied from one to eight. Prompts to count the tones were presented 1.5 s *after* the presentation of critical information and during the presentation of filler narrative information.

After the narrative presentation, all participants completed the final cued recall test. Thus, participants in the interim test group completed this cued recall test once before and once after the narrative. Participants in the standard misinformation group completed the test only after the narrative. All participants were allowed to withhold responses if they could not remember an answer. Following the final cued recall test, participants completed the discrepancy recollection task, and then were debriefed and thanked for their participation.

Pilot testing. To ensure that the secondary task manipulation did not disrupt encoding of narrative details, ten undergraduates from Tufts University served as pilot participants. Participants first listened to the recording of the narrative. Presence of the tone counting task was manipulated within subjects. On half of the critical trials, participants engaged in the secondary tone counting task. A sequence of tones was presented at the offset of critical detail presentation. On the other half of the critical trials, critical details were processed in the absence of the secondary task. After a five minute retention interval, participants completed a cued recall task. After the cued recall task, they completed a two-alternative forced choice recognition task. Participants did not differ in recall accuracy when tones were present ($M = 0.47$) compared to when they were not present ($M = 0.44$), $t(9) = 0.57$, $p > 1$. Participants did not differ in recognition accuracy when tones were present ($M = 0.72$) compared to when they were not present ($M = 0.75$), $t(9) = 0.41$, $p > 1$. These results suggest that the secondary task did not impair encoding of critical details.

Results

Secondary task performance

Participants' performance on the tone counting task is important because it demonstrates successful engagement in the task. In turn, successful engagement in the task suggests that fewer attentional resources would be available to engage in elaboration of critical narrative details. An independent *t*-test compared mean accuracy on the task between the Standard ($M = 0.87$, $SD = 0.27$) and Interim ($M = 0.90$, $SD = 0.20$) test groups. Importantly, the groups performed equally well on the task, $t\,(78) = 0.57$, $p = 0.57$).

Memory performance

Where appropriate, all pairwise comparisons used a Bonferroni correction. During the initial recall test, 0.62 of participants' responses were accurate and 0.06 were characterized as spontaneous misinformation production. Confidence ratings on the final test were not relevant to the hypotheses in this study so were not included in the analyses.

Accurate video recall on the final test. Table 1 presents the accurate video recall probabilities on the final test. A 2 (Testing Group: Standard, Interim) × 2 (Attention: Full, Divided) × 3 (Item type: Consistent, Neutral, Misleading) mixed design analysis of variance (ANOVA) was conducted on final test accurate recall. For the omnibus test, main effects are reported first, followed by interactions. A main effect of Item Type was significant, $F\,(2,312) = 90.91$, $p < 0.001$, $\eta_p^2 = 0.37$. Participants were significantly more accurate on consistent trials ($M = 0.76$) compared to neutral trials ($M = 0.58$), $t\,(158) = 9.90$, $p < 0.01$, $d = 0.78$. In addition, participants were more accurate on neutral trials as compared to misleading trials ($M = 0.49$), $t\,(158) = 4.25$, $p < 0.01$, $d = 0.36$. We found a marginal main effect of Attention, $F\,(1,156) = 3.17$, $p = 0.08$. The main effect of Testing was not significant, $F\,(1,156) = 1.74$, $p = 0.19$.

The ANOVA revealed a significant Item Type by Testing Group interaction, $F\,(2,312) = 6.89$, $p < 0.01$, $\eta_p^2 = 0.04$. The interaction was driven by the difference between testing groups on consistent trials. Interim test participants were more accurate ($M = 0.81$) than standard participants ($M = 0.71$) on these trials, $t\,(158) = 3.07$, $p < 0.05$, $d = 0.49$. Comparisons on neutral, $t\,(158) = 1.6$, $p = 0.11$ and misleading trials, $t\,(158) = 1.3$, $p = 1.0$, were not significant. A Testing Group by Attention interaction was significant, $F\,(1,156) = 9.37$, $p < 0.05$, $\eta_p^2 = 0.06$. In order to tease apart the effect of our attention manipulation on each testing group, we ran two independent samples *t*-tests on mean final test accuracy, collapsing across item type. *T*-tests were corrected for alpha inflation using the Bonferroni method. We found that Attention did not impact final test accuracy for participants in the standard misinformation group, $t\,(78) = 0.81$, $p = 0.42$. However, the attention manipulation did influence performance in the interim test group, $t\,(78) = 3.98$, $p < 0.01$, $d = 0.99$. Interim test participants were more accurate if they encoded the narrative under divided attention constraints ($M = 0.69$) as compared to under full attention ($M = 0.57$).

Table 1
Mean proportion of video details recalled correctly on the final test in Experiment 1 as a function of item type, testing group, and attention (standard error in parentheses).

	Consistent	Neutral	Misleading
Standard/Full	0.75 (0.03)	0.56 (0.03)	0.52 (0.04)
Interim/Full	0.78 (0.03)	0.55 (0.03)	0.38 (0.04)
Standard/Divided	0.68 (0.03)	0.55 (0.03)	0.52 (0.04)
Interim/Divided	0.84 (0.03)	0.66 (0.03)	0.56 (0.04)

Table 2
Mean proportion of commission errors (misleading intrusions) on the final test in Experiment 1 as a function of item type, testing group, and attention (standard error in parentheses).

	Consistent	Neutral	Misleading
Standard/Full	0.03 (0.05)	0.08 (0.08)	0.25 (0.19)
Interim/Full	0.04 (0.07)	0.09 (0.10)	0.44 (0.25)
Standard/Divided	0.05 (0.07)	0.08 (0.08)	0.25 (0.21)
Interim/Divided	0.04 (0.07)	0.09 (0.10)	0.30 (0.27)

Misinformation errors of commission on the final test. Table 2 presents the misinformation errors of commission on the final test. Paralleling our analysis on accurate recall of video details, we performed a 2 (Testing Group: Standard, Interim) × 2 (Attention: Full, Divided) × 3 (Item type: Consistent, Neutral, Misleading) mixed design analysis of variance (ANOVA) on misleading errors of commission. A main effect of Item Type was significant, $F\,(2,312) = 136.20$, $p < 0.001$, $\eta_p^2 = 0.47$. Participants were far more likely to incorrectly report misleading details ($M = 0.33$) on the final test if those details were presented in the narrative, as compared to if consistent ($M = 0.04$) or neutral details ($M = 0.07$) were presented in the narrative, (misleading vs. consistent: $t\,(120) = 14.94$, $p < 0.001$, $d = 1.93$; misleading vs. neutral: $t\,(120) = 13.07$, $p < 0.001$, $d = 1.67$). We found a main effect for Testing, $F\,(1,156) = 12.76$, $p = 0.001$, $\eta_p^2 = 0.08$. Participants in the interim test group ($M = 0.37$) were more likely to produce misleading details than participants in the standard group ($M = 0.25$). The main effect of Attention was not significant, $F\,(1,156) = 2.41$, $p = 0.12$.

The ANOVA revealed a significant Item Type by Testing Group interaction, $F\,(2,312) = 7.36$, $p < 0.001$, $\eta_p^2 = 0.04$. The interaction was driven by the difference between testing groups on misleading trials. Interim test participants were likely to make errors of commission ($M = 0.37$) than standard participants ($M = 0.25$) on these trials, $t\,(158) = 3.28$, $p < 0.01$, $d = 0.50$. Comparisons on neutral, $t\,(158) = 0.83$, $p = 0.40$, and consistent trials, $t\,(158) = 0.48$, $p = 0.62$, were not significant. A Testing Group by Attention interaction was significant, $F\,(1,156) = 4.80$, $p < 0.05$, $\eta_p^2 = 0.03$. In order to tease apart the effect of our attention manipulation on each testing group, we ran two independent samples *t*-tests on mean errors of commission, collapsing across item type. *T*-tests were corrected for alpha inflation using the Bonferroni method. We found that Attention did not impact errors of commission for participants in the Standard misinformation group, $t\,(78) = 0.51$, $p = 0.61$. However, the attention manipulation did influence performance in the Interim test group, $t\,(78) = 2.39$, $p < 0.01$, $d = 0.59$. Interim test participants were less likely to produce misleading errors of commission if they encoded the narrative under divided attention constraints ($M = 0.15$) as compared to under full attention ($M = 0.19$). Finally, we found an Item Type by Attention interaction, $F\,(2,312) = 3.15$, $p < 0.05$, $\eta_p^2 = 0.02$. As demonstrated by Table 2, dividing attention had the largest impact on participants in the Interim test group, specifically on misleading trials. However, comparisons using a Bonferroni correction did not result in a significant effect, $t\,(158) = 1.82$, $p = 0.07$.

Discrepancy recollection

Mean discrepancy recollection is reported in Table 3. A 2 (Testing Group: Standard, Interim) × 2 (Attention: Full, Divided) between groups ANOVA assessed participants' recollection of discrepancies between details presented in the video and narrative on misleading trials. A main effect of Testing Group was found, $F\,(1,156) = 4.67$, $p < 0.05$, $\eta_p^2 = 0.03$. The interim test group accurately remembered discrepancies on 60% of trials, while the standard group remembered discrepancies on 53% of trials. No other effects were significant, $Fs < 1$.

Table 3
Mean proportion of recollected discrepancies between video and narrative reported in Experiments 1 and 2 as a function of item type, testing group, and attention (standard error in parentheses).

	Consistent	Neutral	Misleading
Experiment 1			
Standard/Full	0.26 (0.03)	0.33 (0.04)	0.56 (0.03)
Interim/Full	0.23 (0.03)	0.27 (0.04)	0.59 (0.03)
Standard/Divided	0.29 (0.03)	0.36 (0.04)	0.51 (0.03)
Interim/Divided	0.20 (0.03)	0.27 (0.04)	0.61 (0.03)
Experiment 2			
Standard/Full	0.22 (0.03)	0.36 (0.04)	0.60 (0.03)
Interim/Full	0.18 (0.03)	0.43 (0.06)	0.65 (0.04)
Standard/Divided	0.23 (0.03)	0.41 (0.04)	0.44 (0.04)
Interim/Divided	0.18 (0.03)	0.39 (0.05)	0.58 (0.04)

Experiment 1 discussion

Experiment 1 examined the hypothesis that interim testing changes how subsequently presented misinformation is processed. In support of this hypothesis, when a secondary task disrupted additional processing of misleading post-event details, RES misleading errors of commission were reduced. Importantly, participants in both the interim test and standard misinformation groups demonstrated *baseline* misinformation production effects. That is, both groups produced misleading post-event details on the final test of memory. However, the presence of the secondary task eliminated the *enhanced* errors of commission typically observed after interim testing. When the narrative was presented in conjunction with the secondary task, there was no difference in misinformation production errors between the standard and interim testing groups. These results suggest that the secondary task did disrupt the preferential processing of narrative details that has been hypothesized to be dependent on prior testing. When preferential processing was minimized, so were RES errors of commission. Similarly, interim test participants were most *accurate* on the final test when the narrative was encoded under divided attention. This is likely because these participants did not elaborately process the misleading narrative information that would typically reduce memory accuracy. The recollection of discrepancy analysis found that participants who took an interim test were better able to remember differences between original and post-event information than those who did not take that test. This may indicate that memory for the original event coexists with memory for post-event information (cf., Gordon & Shapiro, 2012), and that testing may promote segregation between the original and post-event information (c.f., Wahlheim, 2015).

The findings from Experiment 1 suggest that elaborative processing of misleading details may be necessary for RES, as measured by increased misinformation errors of commission, to emerge. After taking a test, participants are cued to misleading narrative details that are relevant to the initial test questions (Gordon & Thomas, 2014; Gordon et al., 2015). This test-enhanced attention to misleading narrative details may increase the temporary accessibility of those details in memory, and impact how fluently those details are retrieved from memory (c.f., Thomas et al., 2010). In addition to attention impacting encoding of narrative details, we propose that interim testing may also impact the segregation of the original event from the post-event synopsis, as suggested by the discrepancy recollection data.

Experiment 2

Experiment 2 allowed us to examine both encoding and retrieval based explanations of how forward effects may manifest in RES studies. Experiment 2 had two primary aims. First, we tested whether interim testing potentiated learning of post-event narrative material by directly assessing memory for the narrative. Second, we examined potential factors underlying test-enhanced learning. If testing changes how subsequent narrative material is processed, and ultimately learned, then disruption of that processing should reduce test-enhanced learning of new narrative material. However, if testing also promotes contextual event segregation, then we would expect to see better memory for the narrative and more accurate discrepancy recollection for participants who take an interim test, even when elaborative processing of the narrative is minimized.

As in Experiment 1, the primary manipulation in Experiment 2 was the introduction of a secondary task to reduce test-potentiated elaboration after encoding. In addition, Experiment 2 directly examined memory for the narrative on the final test, or "List 2". Discrepancy recollection data was also collected in this experiment.

Method

Participants

A new sample of fifty-two male and 107 female undergraduates from Tufts University, with a mean age of 19.0 (SD = 2.5), participated and were compensated either with course credit or $15. One participant did not report gender. Of the total 160 participants, 58% identified as White, 17% as Asian, 5% as Black, 6% as Hispanic or Latino, 8% as Multi-racial, and 6% as another race not listed. All participants spoke English as the primary language, or indicated proficiency in English. All participants provided written informed consent prior to beginning the experiment. As with Experiment 1, participants were equally distributed across, and randomly assigned to four groups.

Materials and procedure

The design, materials, and procedure were identical to Experiment 1 with one change. On the final memory test, participants were instructed to report what they remembered learning in the narrative only.

Results

Secondary task performance

An independent t-test compared mean accuracy on the task between the Standard (M = 0.87, SD = 0.26) and Interim (M = 0.89, SD = 0.23) test groups. Importantly, the groups performed equally well on the task, t (78) = 0.411, p = 0.68.

Memory performance

During the initial recall test, 0.63 of participants' responses were accurate and 0.07 produced misinformation spontaneously.

Final test narrative recall. Mean narrative recall accuracy is reported in Table 4. On the final test, participants were instructed to report information they learned in the narrative only. For consistent trials, this was a detail originally presented in the video and thus *repeated* information in the narrative. For misleading trials, this was a detail not presented in the video and thus *new* information in the narrative. In light of potential repetition effects, these trial types fundamentally differed. We examined recall of repeated (consistent) and new (misleading) details separately. Neutral trials did not provide specific details in the narrative so were excluded from analysis. Follow up tests for each analysis used a Bonferroni correction unless otherwise stated.

Consistent trials. A 2 (Testing Group: Standard, Interim) × 2 (Attention: Full, Divided) between groups ANOVA examined recall of

Table 4
Mean proportion of narrative details recalled correctly on the final test in Experiment 2 as a function of item type, testing group, and attention (standard error in parentheses).

	Consistent	Misleading
Standard/Full	0.85 (0.03)	0.53 (0.04)
Interim/Full	0.87 (0.03)	0.82 (0.04)
Standard/Divided	0.73 (0.03)	0.38 (0.04)
Interim/Divided	0.88 (0.03)	0.73 (0.04)

details that were presented in the video and repeated in the narrative. A main effect of Testing Group was revealed, $F(1,156) = 10.17$, $p < 0.01$, $\eta_p^2 = 0.06$. Participants in the Interim test group ($M = 0.88$) recalled more consistent details from the narrative than participants in the standard group ($M = 0.79$). A main effect of Attention was also revealed, $F(1,156) = 4.51$, $p < 0.05$, $\eta_p^2 = 0.03$. Participants in the full attention group ($M = 0.86$) recalled more consistent details from the narrative than participants in the divided attention group ($M = 0.81$). Finally, the Testing Group by Attention interaction was significant, $F(1,156) = 6.11$, $p < 0.05$, $\eta_p^2 = 0.04$. In order to tease apart the effect of the attention manipulation on each testing group, we ran two independent samples *t*-tests using a Bonferroni correction. In the Interim Test group, there was no difference in consistent narrative recall under full ($M = 0.87$) or divided ($M = 0.88$) attention, $t(78) = 0.27$, $p = 0.79$. However in the Standard Test group, dividing attention during narrative encoding impaired consistent item recall ($M = 0.73$) compared to the full attention condition ($M = 0.85$), $t(78) = 2.97$, $p < 0.01$, $d = 0.67$.

Misleading trials. A 2 (Testing Group: Standard, Interim) × 2 (Attention: Full, Divided) between groups ANOVA examined recall of misleading details that were only presented in the context of the narrative. A main effect of Testing Group was revealed, $F(1,156) = 79.48$, $p < 0.001$, $\eta_p^2 = 0.34$. Participants in the Interim test group ($M = 0.78$) recalled more misleading details from the narrative than participant in the standard group ($M = 0.45$). A main effect of Attention was also revealed, $F(1,156) = 11.64$, $p < 0.001$, $\eta_p^2 = 0.07$. Participants in the full attention manipulation ($M = 0.68$) recalled more misleading details from the narrative than participants in the divided attention manipulation ($M = 0.55$). The interaction effect was not significant, $F < 1$.

Neutral trial video intrusions. As participants were instructed to report only what they remembered from the narrative on the final test, a correct answer was an omitted answer. As to be expected, participants rarely spontaneously reported misleading details on these trials (see Table 5). However, participants did sometimes erroneously report that a video detail was re-presented in the narrative on neutral trials. A 2 (Testing Group: Standard, Interim) × 2 (Attention: Full, Divided) between groups ANOVA assessed video intrusions reported on neutral trials. A main effect of Testing Group was found, $F(1,156) = 23.59$, $p < 0.001$, $\eta_p^2 = 0.13$. Participants in the interim test group ($M = 0.32$) reported fewer video intrusions than participants in the standard group ($M = 0.50$). Neither the main effect of Attention nor the interaction were significant, $Fs < 1$. Full means are reported in Table 5.

Table 5
Mean proportion of video and misleading detail intrusions on neutral trials in Experiment 2 as a function of item type, testing group, and attention (standard error in parentheses).

	Video detail	Misleading detail
Standard/Full	0.51 (0.04)	0.06 (0.01)
Interim/Full	0.32 (0.04)	0.07 (0.01)
Standard/Divided	0.49 (0.04)	0.09 (0.01)
Interim/Divided	0.33 (0.04)	0.06 (0.01)

Discrepancy recollection

A 2 (Testing Group: Standard, Interim) × 2 (Attention: Full, Divided) between groups ANOVA assessed participants' likelihood of recollecting a discrepancy between details presented in the video and narrative on misleading trials. A main effect of Testing Group was found, $F(1,156) = 6.46$, $p < 0.05$, $\eta_p^2 = 0.04$. Participants in the Interim test group ($M = 0.61$) were more likely to accurately recollect a discrepancy than participants in the standard group ($M = 0.53$). A main effect of Attention was also revealed, $F(1,156) = 12.64$, $p = 0.001$, $\eta_p^2 = 0.08$. Participants in the full attention group ($M = 0.63$) were more likely to accurately recollect a discrepancy than participants in the divided attention manipulation ($M = 0.51$). The interaction was not significant, $F < 1$. Means are reported in Table 3.

Experiment 2 discussion

In Experiment 2, interim testing enhanced learning of narrative information. Overall, the participants who took an interim test demonstrated greater recall of both consistent and misleading narrative details as compared to participants who did not take an interim test. As learning of the post-event narrative has never been directly measured in the RES paradigm, this is important evidence to support the view that retrieval enhanced suggestibility, as evidenced by test-related increases in misleading errors of commission, may be understood as *test-potentiated learning*, and may be understood within the context of both encoding and retrieval based explanations for the phenomenon (cf., Pastötter & Bäuml, 2014).

In addition to furthering our understanding of RES errors of commission, the results of Experiment 2 contribute to our knowledge of the mechanisms that may underlie the forward effects of testing. Considering first the analysis on misleading trials, the only situation in which new information was presented after testing, the results clearly demonstrated that testing facilitated new learning. However, this test-enhanced new learning occurred even under divided attention, where the ability to preferentially process narrative details was minimized. Considering next the analysis on consistent trials, when participants were directly asked to recall information from the narrative the secondary distractor task negatively impacted recall of those details in the standard misinformation group, but had *no impact* in the interim testing group. Taken together, these results suggest that whereas preferential processing, as operationalized by additional time to think about and elaborately encode post-event details, may be important for RES effects as measured by misinformation errors of commission (Experiment 1), it was less relevant when post-event learning was directly assessed (Experiment 2). In other words, dividing attention during narrative encoding in Experiment 1 reduced the degree of misleading errors of commission typically observed in the interim test group, but the same manipulation in Experiment 2 did *not* reduce learning of the narrative. A proposed explanation for this set of findings is presented in the General Discussion.

Finally, replicating Experiment 1, testing prior to narrative presentation increased participants' likelihood of accurately reporting discrepancies between the original video and post-event narrative, even when attention was divided during narrative encoding. Further, interim test participants were less likely than the standard group to erroneously report video details on neutral trials where the video detail was not repeated in the narrative. Taken together, these findings provide additional evidence that testing changes

how subsequently presented details are processed and stored in memory, perhaps by facilitating segregation between learning episodes.

General discussion

In two experiments we tested both encoding- and retrieval-based explanations for retrieval enhanced suggestibility. In Experiment 1, we found evidence that testing influenced the way subsequent information was processed. When the ability to allocate extra attention to test-relevant post-event details was minimized by the presence of a secondary task, participants who took an interim test were no more likely to incorrectly report misleading details from the narrative than participants who did not take an interim test. In Experiment 2, we found that interim testing enhanced learning of new misleading details presented in the context of a post-event narrative. However, while disrupting post-encoding processing of narrative details reduced retrieval enhanced suggestibility in Experiment 1, the same manipulation did not have a comparable impact on test-potentiated learning that was directly measured in Experiment 2. Intervening testing promoted learning of new narrative details even when elaborative processing of those details was minimized by the secondary task. Finally, in both experiments participants who were initially tested were better able to accurately recollect a discrepancy between the video and narrative on misleading trials. This general discussion first addresses the pattern of findings across experiments. The remainder of the discussion focuses on the implications of the findings for RES and also broader theories of test enhanced learning effects.

Differentiating between suggestibility and learning

A primary goal of this RES study was to explore how interim testing impacts narrative processing, and how changes to narrative processing may increase retrieval fluency of misleading narrative details (misleading errors of commission). We proposed that interim testing changes how individuals processing narrative details. This change in processing enhances learning of misleading narrative details, and the likelihood of erroneously producing those misleading details on later memory tests. If this argument is valid, then it is unclear as to why experimentally controlling attentional processes during presentation of the narrative decreased inaccurate reporting of misleading details (Experiment 1), but did not decrease learning of those misleading details (Experiment 2).

When making sense of these seemingly divergent results, it is important to consider that the nature of the final memory tests fundamentally differed between experiments. In Experiment 1, participants were instructed to respond with the original event detail only. In Experiment 2, they were instructed to respond with information learned from the narrative. In the case of Experiment 1, we contend that producing misleading narrative details reflects memory errors that may be further influenced by retrieval fluency mechanisms. That is, when relevant test questions precede encoding of misleading narrative details, the additional attention allocated to encoding those details increase their temporary accessibility in memory. In turn, this influences the ease with which misleading details are retrieved on the final test (c.f., Thomas et al., 2010). We found that disrupting any additional attention given to misleading details during the narrative presentation successfully decreased misleading errors of commission in the interim test group.

Although we argue that the results of Experiment 1 suggest that disruption to preferential processing due to the secondary task reduced retrieval fluency associated with misleading details thereby eliminating the RES misinformation production effect, an alternative explanation exists. The disruption of preferential processing of the narrative may simply have reduced learning of misleading details. Experiment 1 did not differentiate between these two possibilities; however Experiment 2 suggests that a disruption of preferential processing, as operationalized by the introduction of a secondary task, may not impact test-enhanced learning, when narrative memory was directly assessed. That is, participants who took an interim test performed better when directly tested on memory for the narrative as compared to participants who did not take an interim test.

While the final memory test in Experiment 1 was not set up to directly test test-enhanced narrative learning as it was in Experiment 2, we contend that test-enhanced learning of misleading narrative details likely occurred even when attention was divided during narrative encoding. There are two pieces of evidence to support this assumption. First, on the discrepancy recollection task in Experiment 1, interim test participants were better able to accurately identify that two pieces on conflicting information were presented on misleading trials compared to participants who did not take an interim test, even when their attention was divided. This indirectly indicates test-enhanced learning. Second, as previously mentioned, the results of Experiment 2 suggest that interim testing enhanced learning of misleading narrative details even when attention to the narrative was divided. Taken together, the present experiments suggest that both retrieval fluency as influenced by changes in narrative processing, as well as event segmentation, impact final test performance depending on the nature of the final test.

How testing increases misinformation susceptibility

The present findings support the perspective taken by Gordon, Thomas, and colleagues (Gordon & Thomas, 2014; Gordon et al., 2015; Thomas et al., 2010), who suggest that testing changes the way subsequent post-event information, notably inaccurate information, is processed. This change in processing impacts susceptibility to misleading information. As prior research has demonstrated, testing prior to post-event information affects reading time associated with sentences relevant to previously tested details, particularly those that contradict initial memory reports (Gordon & Thomas, 2014; Gordon et al., 2015). In the present study, we controlled test-related differences in processing time by pairing an experimenter-paced aural narrative with a secondary distractor task. The secondary task was designed to minimize the additional time individuals may spend thinking about critical details that conflicted with the original event. Processing the narrative in the context of the secondary task did not eliminate the typical misinformation effect. That is, participants did encode and sometimes inappropriately produced misinformation on the final test of memory. However, processing of the narrative in the context of the secondary task did eliminate the *enhanced* production effect. Interim testing did not lead to greater production of misleading details as compared to the standard misinformation group who did not take this initial test.

It is our view that changes in processing associated with misleading narrative details, as a result of interim testing, influences the temporary accessibility of those details in memory. An increase in temporary accessibility may then have influenced the ease with which misleading narrative details came to mind, biasing responding on a final memory test (cf., Baddeley, 1982; Jacoby & Dallas, 1981). In support of this explanation, concurrent findings from our research team show that when temporary accessibility of post-event information was reduced by increasing the retention interval between narrative processing and final testing, RES was eliminated (Thomas, Gordon, Cernasov, & Bulevich, submitted for publication). Further, when participants were encouraged to rely

on recollective experiences as opposed to temporary accessibility, RES effects were also eliminated (Thomas et al., 2010). Finally, in the present Experiment 1, inhibiting interim test participants' ability to spend additional time processing misleading narrative details may have effectively neutralized any increases in misinformation accessibility above and beyond what occurs in the absence of interim testing. As in previous studies, dampening of retrieval fluency translated to reduced RES production effects.

Test-enhanced learning

Experiment 2 revealed that after testing, individuals learned subsequently presented information better. Interim test participants were more likely to recall new, misleading narrative details on a final test of memory compared to participants who did not take an interim test. This experiment was important for two reasons. First, it is the first straightforward demonstration of test-enhanced learning in the RES paradigm. Second, directly measuring post-event learning afforded a novel way to examine theories of test-enhanced learning. Theories of test potentiated learning, or forward effects of testing, can be divided into two basic categories: retrieval and encoding explanations. We predicted an encoding explanation of test-enhanced learning. Encoding explanations propose that testing changes participants' encoding strategy, enhancing attention to and encoding of post-test information informative (e.g., Reynolds & Anderson, 1982; Wissman et al., 2011). We posited that if increased attention was important to RES, and RES is an instance of test-potentiated learning, then attention should also be important to test-potentiated learning.

Interestingly, the ability to spend extra time attending to narrative details was not necessary for the forward effect of testing. Even when the ability to preferentially process new details was disrupted, test-enhanced learning of those details occurred. While there is possibility that our manipulation did not fully disrupt additional processing of narrative details in Experiment 2, the findings from Experiment 1 suggest otherwise. The results from Experiment 2 join a general consensus that neither backward (enhanced memory for tested information) nor forward (enhanced learning of new information) testing effects can be solely explained by an attention mechanism. 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 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) 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 of the passage on a final test compared to participants who were given the passage with important details in highlighting typeface. Essentially, taking a test led to greater learning than comparative attention manipulations.

In contrast to an encoding account, retrieval explanations of testing effects assume that testing encourages differentiation between learning episodes. When initially learned information is tested, it is strengthened (e.g., Roediger & Karpicke, 2006). This process counteracts proactive interference, or interference from originally learned information, by segregating source information from the original learning episode and subsequent learning episodes (e.g., Bäuml & Kliegl, 2013; Pastötter et al., 2011; Szpunar et al., 2008; Weinstein et al., 2014). On later tests, more information can be recalled from subsequent learning episodes because individuals can use source information as a retrieval cue, and are better equipped to accurately monitor sources. The resulting memory performance manifests as test-potentiated learning or forward effects. Importantly, the pattern of results across the present experiments supports a retrieval-based explanation of test-potentiated learning of new, misleading information, and extends the investigation of test-potentiated learning to event memory paradigms. In Experiment 2, participants who took an initial test recalled more new information from the narrative than participants who were not tested. Further, both experiments revealed that the interim test participants recalled more discrepancies between video and narrative content than the standard group. Accurate discrepancy recollection suggests source segregation.

Conclusions

This study was built upon the premise that RES is an indirect example of test potentiated learning. We found that testing promoted learning of new details in the present paradigm, and those new details were misleading details. One way to increase the likelihood of retrieving newly learned misleading details on later memory tests is to change how individuals process misleading post-event details. The allocation of processing resources fundamentally changes if testing is introduced prior to processing. We found that while preferential processing of post-event information appears necessary for retrieval enhanced suggestibility, even when post-event information processing was minimized retrieval enhanced new learning. This suggests that while changes to encoding processes may be necessary for enhanced suggestibility, it is not the primary mechanism underlying test-enhanced learning. That is, retrieval enhanced suggestibility effects may be mediated by retrieval fluency biases that arise from changes in narrative encoding processes.

The present research takes an important step toward understanding how post-event retrieval can influence eyewitness suggestibility. We argue that elaboration in encoding may impact temporary accessibility of misleading details leading to a retrieval fluency bias during a final memory test. This has important implications for eyewitness memory. When an eyewitness's attention is drawn to inaccurate details about an event, perhaps after answering questions at the scene of the crime, eyewitnesses may be more likely to report these inaccuracies when they later recount the crime. However, if appropriate retrieval cues are presented that encourage careful source discrimination, eyewitnesses override this bias and successfully retrieve the correct event details.

Author note

All work described in this manuscript was carried out in accordance with the Code of Ethics for use of human subjects described in the Declaration of Helsinki.

Funding

This work was funded by the American Psychological Association and Tufts University.

References

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

Aslan, A., & Bäuml, K.-H. T. (2016). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science, 19*, 992–998.

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

Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language, 68*, 39–53.

Bezard, R., Bérard, H., Cabaud, P., & Dassin, J. (1955). *Rififi [motion picture]*. France: Pathé Consortium Cinéma.

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, & Cognition, 37*, 1547–1552.

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

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.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences, USA, 103*, 11778–11783.

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.

Gordon, L. T., & Shapiro, A. M. (2012). Priming correct information reduces the misinformation effect. *Memory & Cognition, 40*, 717–726.

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

Gordon, L. T., Thomas, A. K., & Bulevich, J. B. (2015). Looking for answers in all the wrong places: How testing facilitates learning of misinformation. *Journal of Memory and Language, 83*, 140–151.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40*, 505–513.

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 290–296.

Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology, 8*, 200–224.

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., Wahlheim, C. N., & Kelley, C. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41*, 1282–1297.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*, 469–486.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 989–998.

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, & Cognition, 37*, 287–297.

Putnam, A. L., Wahlheim, C. N., & Jacoby, L. L. (2014). Memory for flip-flopping: Detection and recollection of political contradictions. *Memory & Cognition, 42*, 1198–1210.

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

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.

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

Sahakyan, L., & Hendricks, H. E. (2012). Context change and retrieval difficulty in the list- before-last paradigm. *Memory & Cognition, 40*, 844–860.

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.

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.

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., Gordon, L. T., Cernasov, P., & Bulevich, J. B. (submitted for publication). More than changes in encoding: How interim testing promotes learning of misinformation.

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.

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition, 43*, 27–38.

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive reminding in proactive effects of memory. *Memory & Cognition, 41*, 1–15.

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1039–1048.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*, 449–455.

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, 20*, 81–93.

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.