9-2013

# Testing Increases Suggestibility for Narrative-based Misinformation But Reduces Suggestibility for Question-based Misinformation

Jessica A. LaPaglia
*Morningside College*

Jason C.K. Chan
*Iowa State University*, ckchan@iastate.edu

# Testing Increases Suggestibility for Narrative-based Misinformation But Reduces Suggestibility for Question-based Misinformation

**Abstract**
A number of recent studies have found that recalling details of an event following its occurrence can increase people's suggestibility to later presented misinformation. However, several other studies have reported the opposite result, whereby earlier retrieval can reduce subsequent eyewitness suggestibility. In the present study, we investigated whether differences in the way misinformation is presented can modulate the effects of testing on suggestibility. Participants watched a video of a robbery and some were questioned about the event immediately afterwards. Later, participants were exposed to misinformation in a narrative (Experiment 1) or in questions (Experiment 2). Consistent with previous studies, we found that testing increased suggestibility when misinformation was presented via a narrative. Remarkably, when misinformation was presented in questions, testing decreased suggestibility. Copyright © 2013 John Wiley & Sons, Ltd.

**Keywords**
misinformation effect, misleading questions, eyewitness memory, retrieval-enhanced suggestibility



**Comments**
This is the accepted version of the following article: Special Issue: Memory Formation and Suggestibility in the Legal Process Volume 31, Issue 5, pages 593–606, September/October 2013, which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1002/bsl.2090/full.

This is the accepted version of the following article: Special Issue: Memory Formation and Suggestibility in the Legal Process Volume 31, Issue 5, pages 593–606, September/October 2013, which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1002/bsl.2090/full.

# Testing Increases Suggestibility for Narrative-based Misinformation But Reduces Suggestibility for Question-based Misinformation

Jessica A. LaPaglia

*Morningside College*

Jason C. K. Chan

*Iowa State University*

## Author Note

These data were collected as part of the dissertation of the first author. We thank the following research assistants for their help with data collection: Jennifer Dillon, Kelsi Dodd, and Jennifer Piatak.

Correspondence concerning this article should be addressed to Jessica LaPaglia, Department of Psychology, Morningside College, 1501 Morningside Ave, Sioux City, IA 51106, USA.

E-mail: lapagliaj@morningside.edu

Abstract

A number of recent studies have found that recalling details of an event following its occurrence can increase people's suggestibility to later presented misinformation (e.g., Chan, Thomas, & Bulevich, 2009). However, several other studies have reported the opposite result, whereby earlier retrieval can reduce subsequent eyewitness suggestibility (e.g., Pansky & Tenenboim, 2011). In the present study, we investigated whether differences in the way misinformation is presented can modulate the effects of testing on suggestibility. Participants watched a video of a robbery and some were questioned about the event immediately afterwards. Later, participants were exposed to misinformation in a narrative (Experiment 1) or in questions (Experiment 2). Consistent with previous studies, we found that testing increased suggestibility when misinformation was presented via a narrative. Remarkably, when misinformation was presented in questions, testing decreased suggestibility.

*Keywords*: misinformation effect, misleading questions, eyewitness memory, retrieval-enhanced suggestibility

# Testing Increases Suggestibility for Narrative-based Misinformation But Reduces Suggestibility for Question-based Misinformation

In a recent New Jersey ruling, there was a reformulation of juror instructions aimed at clarifying how eyewitness testimony should be evaluated (Weiser, 2012). The new instructions educate jurors on findings from decades of research demonstrating the fallibility of eyewitness memory. Indeed, memory for an event or perpetrator can be drastically altered when witnesses are exposed to misleading postevent information (the misinformation effect; for a review see Loftus, 2005). More relevant to the present research is a counterintuitive finding that has been reported in a series of studies. Namely, that recalling the details of an event can increase eyewitness suggestibility to later presented misinformation (e.g., Chan, Thomas, & Bulevich, 2009; Chan & LaPaglia, 2011).

This finding, termed retrieval-enhanced suggestibility (RES), is surprising because retrieval often enhances retention of an event (i.e., the testing effect; Roediger & Karpicke, 2006). Given that retrieval practice is such a powerful memory enhancer, the logical prediction is that completing a recall test for a witnessed event soon after its occurrence should enhance retention of that event and thus protect witnesses from later suggestions. Consistent with this idea, having a better memory for a witnessed event is often associated with lower susceptibility to misinformation. For example, improved encoding of a witnessed event can reduce the misinformation effect (Lane, 2006; Pezdek & Roe, 1995). Similarly, the impact of misinformation is weaker when the witnessed event is still "fresh in mind" than if forgetting has set in (e.g., Chan & Langley, 2011; Loftus, Miller, & Burns, 1978). However, contrary to the prediction that testing would reduce suggestibility, Chan and colleagues have repeatedly found that testing increased the misinformation effect (e.g., Chan & Langley, 2011; Chan et al., 2009;

Chan, Wilford, & Hughes, 2012). One explanation for this is that completing the initial test paradoxically enhanced learning of the subsequent misinformation, thus increasing its likelihood of being recalled (e.g., Zaragoza & Mitchell, 1996). In line with this hypothesis, testing has been shown to improve subsequent encoding of new materials (Wissman, Rawson, & Pyc, 2011; Szpunar, McDermott, & Roediger, 2008; Tulving & Watkins, 1974).

Applied to the RES effect, the initial recall test may draw attention to specific parts of the narrative, thus enhancing encoding of the misinformation presented in the narrative. For example, in Chan et al.'s (2009) experiments, participants were asked about the weapon used by a terrorist to knock out a flight attendant. Later, participants heard misinformation that the terrorist used a *chloroform rag* when the weapon was actually a *hypodermic syringe*. When participants were asked about the weapon in the initial memory test, it may have inadvertently drawn their attention to this relevant misleading detail in the narrative, thereby enhancing encoding of this misinformation and increasing misinformation recall on the final test. Gordon and Thomas' (2013) recent study lent support to this hypothesis when they found that participants who were initially tested spent more time reading sentences that included misinformation.

Although the RES effect shows that testing can be detrimental to the accuracy of eyewitness memory, several studies have reported the contrary, such that initial testing can protect eyewitnesses from misleading suggestions (e.g., Gabbert, Hope, Fisher, & Jamieson, 2012; LaPaglia & Chan, 2012; Memon, Zaragoza, Clifford, & Kidd, 2010; Pansky & Tenenboim, 2011; Saunders & MacLeod, 2002). Numerous methodological differences between these sets of studies likely contribute to the disparity in the effects of testing on suggestibility. For example, compared to studies demonstrating RES, the studies that showed a testing effect

used different to-be-remembered materials (faces instead of events, LaPaglia & Chan, 2012), types of misinformation suggested (central details instead of peripheral details, Gabbert et al., 2012; self-fabricated misinformation instead of experimenter-provided misinformation, Memon et al., 2010), and type of initial test administered (the Self-Administered Interview instead of a standard recall test, Gabbert et al., 2012), etc. It is beyond the scope of the present study to address the potential contribution of all of these variables, so we focus on one factor that might determine whether initial testing enhances or exacerbates eyewitness suggestibility. Specifically, we examined whether variations in the way misinformation is presented (i.e., in a narrative or misleading questions) modifies the influence of prior testing on suggestibility. This variable is of particular interest because studies that have found initial testing to reduce suggestibility often presented misinformation in questions (Pansky & Tenenboim, 2011; Saunders & MacLeod, 2002), whereas studies that have demonstrated RES always presented misinformation in a narrative.

In an experiment conducted by Saunders and MacLeod (2002), participants read two written narratives describing separate burglaries. After the encoding phase, participants were tested on half of the items from one of the burglaries while all remaining items were not tested. Participants then completed a free recall test for all the studied items before misinformation was introduced via misleading questions. For instance, one narrative described that the burglars stole a *necklace* that was next to the sink. During the misinformation phase, participants responded to the question, "When the burglars stole the *earrings* that were next to the sink in the kitchen, they knocked some items on the floor breaking them. What did they break?" Performance was then measured in a forced-choice recognition test. Unlike studies reporting RES, Saunders and

MacLeod found that participants were less likely to select the misinformation for the tested items than the nontested items.

Three key differences between Saunders and MacLeod's (2002) study and studies that found RES might have contributed to these contrasting findings. First, initial retrieval accuracy was far higher (~89% of questions were recalled correctly) in Saunders and MacLeod's study than was typically observed in studies that reported RES (~55% initial test accuracy). This is important because highly recallable items (i.e., easy items, central items) are more resistant to the RES effect than items that are more difficult to remember (e.g., peripheral items, Chan & Langley, 2011; Wilford, Chan, & Tuhn, 2012). It may be that Saunders and MacLeod's critical items were easily remembered and thus not susceptible to the RES effect. Second, because all participants had completed a free recall test before they were exposed to misinformation, there was no true no-test control condition in this study. Third, and of particular relevance to the current study, the misinformation was presented within misleading questions (as opposed to a narrative). Retrieving information while being exposed to misinformation is different than passively listening to or reading misinformation presented in a narrative, and this may play a key role in determining whether testing would increase or decrease subsequent suggestibility.

In Pansky and Tenenboim's (2011) study, participants viewed a slideshow as their critical event and then took a test over some of the information before they were exposed to misinformation. A final test was given 48 hr later. Pansky and Tenenboim found that participants were less suggestible for the tested items than the nontested items. This study is of particular interest because its methodology is highly similar to the studies reporting RES. In an attempt to explain the inconsistencies between their results and those reported by Chan et al. (2009), Pansky and Tenenboim pointed to four methodological differences. These include the

length of the event video, initial testing manipulated within- vs. between-subjects, a 48 hr vs. 30 min retention interval, and misinformation introduced via questions vs. narrative. Based on current knowledge, the first three differences are likely irrelevant, because the RES effect has been demonstrated across different witnessed event lengths (Chan et al., 2009; Chan et al., 2012), retention intervals (Chan & Langley, 2011), and in both within- and between-subjects designs (Chan & LaPaglia, 2011). We are thus left with the fourth methodological difference – that Pansky and Tenenboim presented their misinformation via written questions whereas studies showing RES presented misinformation in a narrative.

As is apparent from this brief review, initial testing can reduce or increase suggestibility, although the processes underlying the opposite patterns of results are unknown. We suspect that these discrepant findings are related to the way in which misinformation is presented. Here we examine whether variations in how misinformation is delivered (namely, via questions or via a narrative) alter the influence of testing on eyewitness suggestibility.

Several studies have investigated the impact of misinformation presentation method on suggestibility, and the results are mixed. For example, Zaragoza and Lane (1994) found that presenting misinformation in questions increased suggestibility relative to presenting misinformation in a narrative. Similarly, Saunders (2009) reported that misleading questions are more effective at inducing false memories than is a narrative, though this effect was only observed for misinformation regarding central details and not peripheral details. Gobbo (2000) showed that presenting misinformation in questions alone produced a smaller misinformation effect than presenting misinformation in both questions and a narrative (such that the misinformation was repeated). Unfortunately, it is impossible to disentangle the effects of misinformation repetition and presentation method in this study. Moreover, presenting

misinformation in questions that do not presuppose the truth of the misinformation, but rather questions its existence (e.g., I used the Dove soap, was it Dove?), can eliminate the misinformation effect altogether (Lee & Chen, 2012). In sum, if we were to draw a preliminary conclusion based on extant findings, it appears that question-based misinformation might be more effective at inducing false recall than is narrative-based misinformation, though this effect is far from universal. More important for present purposes, however, is that we know virtually nothing about whether testing would alter this tentative relationship.

**The Current Study**

In two experiments, participants watched a video of a bank robbery. Afterwards, participants in the test condition took an initial recall test over the contents of the video. For instance, one critical detail in the present experiments is the amount of time the bank robber gives the police to meet his demands. In the initial test, participants were asked, "How long does the robber give the police to get him the money?" The correct answer to this question is ***two hours***. After a short delay, they either listened to a narrative containing misinformation (Experiment 1) or responded to new questions containing misinformation (Experiment 2). If the misinformation was presented in a narrative, participants heard "The robber demands that the police get him his money and a car in ***one hour***." If the misinformation was presented in a question, participants were asked "In addition to the money, the robber gives the police ***one hour*** to also get him what?" Later, all participants completed a final recall test over the video. The question here was whether an RES effect would remain when misinformation was presented in questions.

We hypothesized that, compared with presenting misinformation in a narrative, presenting misinformation in questions may result in an elimination of the RES effect.

Attempting to answer a question, as opposed to passively listening to a narrative, could encourage participants to engage in more effortful processing (Carpenter, 2009; Pyc & Rawson, 2009). More concretely, research has shown that taking a recall test can sometimes trigger spontaneous retrieval of related information in memory (Chan, 2009). When placed into the present context, answering the misleading question about what the robber demanded (a getaway car) may lead participants to covertly retrieve the related detail (i.e., he needed the car in two hours). This would increase the likelihood that one would detect a discrepancy between one's memory (from either the video event itself or from the response produced during the initial test, or both) and the misinformation (e.g., the robber needed the car in one hour), thereby reducing the influence of the misinformation (Tousignant, Hall, & Loftus, 1986). In fact, based on this logic, it is possible that presenting misinformation in questions might not only eliminate the RES effect, it may reverse it, such that testing would reduce the misinformation effect. Notably, we believe that question-based misinformation would reduce suggestibility relative to narrative-based misinformation only for participants who have received an initial test. Specifically, we believe that participants would be far more likely to notice contradictions between their memory and the misinformation in the question if they have produced an overt response of that very detail during the initial test. A further prediction from this logic is that testing would reduce question-based misinformation only if participants can recall the critical detail correctly during the initial test.

## Experiment 1

In the first experiment, we attempted to extend the RES effect with narrative presentation to a new set of witnessed event materials. Although a number of studies have demonstrated RES, most have used the same 40 min long video of a terrorist attack as the witnessed event

(e.g., Chan et al., 2009; Chan & LaPaglia, 2011). One other published study has demonstrated the effect with a 10 min video event of a museum burglary (Chan et al., 2012). Here we developed a new video featuring a bank robbery, and the length of the video was somewhere between the longer (40 min) video and the shorter (10 min) video. Therefore, finding RES with this new set of materials would further ascertain its generalizability.

## Method

**Participants and Design**. Experiment 1 used a 2 (test vs. no-test) X 2 (postevent information: neutral vs. misled) mixed design. Test vs. no-test was manipulated between subjects. Postevent information was manipulated within subjects. There were 20 participants in each between-subjects condition for a total of 40 participants (23 male). Their mean age was 19.50 (*SD* = 1.18). Participants were undergraduate students from a Midwestern university who participated in this experiment for partial course credit. The experiment consisted of four main phases. First, participants completed the encoding phase in which they viewed the witnessed event video. Second, participants completed the no-test/test phase. Here, they either performed a distractor task or a recall test. Third, participants were exposed to misinformation via an audio narrative. Fourth, participants were administered the final recall test.

**Materials and Procedure.** Participants began by watching a 25 min video from an episode of the television show *Flashpoint*. The video depicted a bank robbery by a disgruntled former employee. Participants were given intentional encoding instructions. Specifically, they were told to pay close attention to the video, including the actions and surrounding environment, because their memory would be tested later. Following the video, participants in the no-test condition played the videogame Tetris for 7 min as a distractor activity, whereas participants in the test condition took a cued recall test over their memory for the video. This test consisted of

14 open-ended, nonleading questions (e.g., “How long does the robber give the police to get him the money?”). Participants were given 30 sec to answer each question and the initial test phase lasted 7 min. They were told to be as accurate as possible and not to guess. No corrective feedback was given.

Once participants completed the initial test or Tetris, they were shown another video to fill a retention interval. The video was a clip from the BBC show *Spooks*. It depicted a terrorist plot against abortion doctors and lasted approximately 20 min. Participants were told that they should pay close attention to the video, but they were not told whether their memory for the video would be tested later.

Following the distractor video, participants were presented with misinformation embedded within an audio narrative via headphones. The audio narrative lasted roughly 6 min. All 14 critical details queried in the initial test were included in the narrative. The critical details were either presented as a misled item or a neutral item. For example, one critical detail was the amount of time the robber gave police to get him the money. If this detail was presented as misinformation, the participants heard that the robber gave the police one hour (when he had in fact given them two). If this detail was presented as a neutral item, participants heard that the robber asked for money, but the amount of time was not specified. Half of the items were misled and the other half were neutral. Whether a detail was misled or not was counterbalanced across participants.

Following a 25 min filled retention interval in which participants completed the Reading Span working memory task (Unsworth, Heitz, Schrock, & Engle, 2005) and played Tetris, they took the final test (which was identical to the initial test). Participants were told to answer the questions based only on their memory for the video. Immediately after each question,

participants were asked to rate their confidence in their response from 1 ("I guessed") to 5 ("I am very sure"). After the final test, participants completed a short demographic questionnaire.

## Results and Discussion

### Initial Test

Responses in the initial and final tests were classified as *correct*, matching the *misinformation* presented later (i.e., spontaneous misinformation recall), *no response* (i.e., a blank response or "I don't know"), or an *other* response (i.e., any response that was incorrect but did not match the misinformation). Correct recall probability was similar to previous RES studies ($M = .64$) and spontaneous misinformation recall probability was, as expected, very low ($M = .07$). See Table 1 for the complete data on the initial test.

### Final Test

We focus our data analysis on the correct and misinformation recall probabilities (see Figure 1). For the sake of completeness, data regarding the "other" and "no response" types are displayed in Table 2.

**Correct Recall.** A 2 (no-test, test) X 2 (postevent information: misled, neutral) ANOVA revealed a significant interaction, $F(1, 38) = 4.14$, $p = .05$ , $\eta_p^2 = .10$. Specifically, the tested participants recalled fewer correct details ($M = .35$) than the nontested participants ($M = .52$) for the misled items, $t(38) = 3.35$, $p = .002$, $d = .80$, but not for neutral items, $t < 1$, $p = .73$. That is, initial testing reduced accurate recall when participants were misled – a reversed testing effect. There was also a significant main effect of postevent information, such that participants recalled more correct details for neutral items ($M = .59$) than for misled items ($M = .44$), $F(1, 38) = 10.51$, $p = .002$ , $\eta_p^2 = .22$. The main effect of initial testing was not significant, $F(1, 38) = 2.83$, $p = .10$, $\eta_p^2 = .07$.

**Misinformation Recall.** A 2 (no-test, test) X 2 (postevent information) ANOVA revealed a significant interaction, $F(1, 38) = 13.72$, $p = .001$, $\eta_p^2 = .27$. Testing nearly doubled the misinformation recall probability for the misled items, $t(38) = 3.35$, $p = .002$, $d = 1.05$ – an RES effect ($M = .47$ for test and $M = .24$ for no-test), but (as expected) not for the neutral items, $t < 1$, $p = .44$. Not surprisingly, there was a significant misinformation effect overall, with greater misinformation recall probabilities for the misled items ($M = .35$) than for the neutral items ($M = .07$), $F(1, 38) = 70.99$, $p < .001$, $\eta_p^2 = .65$. The main effect of testing was also significant, $F(1, 38) = 6.93$, $p = .01$, $\eta_p^2 = .15$.

During testimony, a witness may be prompted to only respond when they are highly confident. Using the confidence data collected in the final test phase, we examined misinformation recall probability only for responses in which participants were highly confident (a confidence rating of 4 or 5—"I am sure" or "I am very sure"). Remarkably, even among these highly confident responses, the RES effect remained intact ($M = .48$ for tested and $M = .25$ for nontested), $t(38) = 2.37$, $p = .02$, $d = .77$. To further investigate the effect of initial testing on later suggestibility, we performed a conditional analysis examining final test accuracy depending on whether the participant successfully recalled an item on the initial test. Not surprisingly, tested participants who were initially incorrect for a given item were more likely to recall the misinformation on the final test ($M = .56$) than otherwise ($M = .43$ for initially correct items), although the difference was not statistically significant, $t(19) = 1.29$, $p = .21$, $d = .37$. However, more surprising is that being able to recall an item correctly during the initial test by no means protected one from the influence of RES, as the misinformation recall probability for these initially-correct items was still far greater than that in the no-test condition ($M = .24$), $t(38) = 3.76$, $p < .001$, $d = 1.19$. Moreover, misinformation had a powerful negative effect on retention

of these initially-correct items, dropping correct recall probability on the final test from .88 (neutral items) to .54 (misled items), $t(19) = 3.29$, $p = .003$.

## Experiment 2

In Experiment 1, we generalized the RES effect to a new set of materials. Experiment 2 included the same four phases as Experiment 1 (i.e., encoding, no-test/test, misinformation introduction, final test), except that the misinformation was embedded in questions instead of a narrative.

### Method

**Participants**. A total of 40 participants (25 male) were tested, with 20 each in the no-test and test conditions, respectively. Their mean age was 19.65 ($SD = 2.09$).

**Materials and Procedure.** The materials and procedure of Experiment 2 were identical to Experiment 1 with the exception of the misinformation phase. During the misinformation phase, participants were presented with 14 questions, each querying a noncritical detail about the video. Seven of the questions included a piece of misinformation (misled items) whereas the remaining questions presented no misleading information (neutral items). Participants were given 30 sec to answer each question. One question that included misinformation was "In addition to the money, the robber gives the police one hour to also get him what?" In this case, like Experiment 1, the misinformation was the time (one hour) the robber gave the police. Note that participants were always asked about the non-critical details during the misinformation phase.

### Results and Discussion

#### Initial Test

Initial and final test data were coded in the same manner as described in Experiment 1. See Table 1 for initial test recall probabilities. Not surprisingly, the correct recall probability was similar to Experiment 1 ($M$ = .58) and spontaneous misinformation recall probability was again low ($M$ = .03).

**Misleading Questions Phase**

In Experiment 1, the misinformation was presented in a narrative that did not require any responses from participants. In Experiment 2, the misinformation was introduced in a memory test of the noncritical details. Responses from the misleading questions were coded as either *correct*, *no response*, or *other* responses. Recall probabilities are presented in Table 3. Correct recall probability was marginally lower for the tested participants ($M$ = .76) than the nontested participants ($M$ = .82), $t(38) = 1.78$, $p$ = .08, $d$ = .55. Tested participants had more "I don't know" and blank responses ($M$ = .08) than nontested participants ($M$ = .04); this difference was again only marginally significant, $t(38) = 1.88$, $p$ = .07, $d$ = .55. These data suggest that completing the initial test might have disrupted recall of related information, a finding known as retrieval-induced forgetting (Anderson, 2003). However, we caution against any strong conclusions here because the difference was small and only marginally significant. We consider the possible cause for this pattern more thoroughly in the General Discussion.

**Final Test**

**Correct Recall.** Correct and misinformation recall probabilities are presented in Figure 2 (see Table 2 for data in the Other and No Response categories). A 2 (test, no-test) X 2 (postevent information) ANOVA showed no significant interaction or main effects of testing and postevent information, $F$s < 2.52, $p$s > .12.

**Misinformation Recall.** As expected, a 2 (test, no-test) X 2 (postevent information) ANOVA revealed a significant misinformation effect with participants reporting more misinformation for the misled items ($M$ = .20) than for the neutral items ($M$ = .02), $F(1, 38) = 27.34$, $p < .001$, $\eta_p^2 = .42$. Far more interesting, however, is that a testing effect, instead of RES, was observed, $F(1, 38) = 12.34$, $p = .001$, $\eta_p^2 = .25$, such that overall misinformation recall probability was higher for the nontested participants ($M$ = .14) than for the tested participants ($M$ = .08). Moreover, these main effects were qualified by a marginally significant interaction, $F(1, 38) = 3.00$, $p = .09$, $\eta_p^2 = .07$, which is driven by the fact that testing substantially reduced misinformation recall probability for the misled items ($M$ = .14 for the test condition and $M$ = .26 for no-test), $t(38) = 2.40$, $p = .02$, $d = .78$, but (unsurprisingly) not for the neutral items, $t(38) = 1.44$, $p = .16$.

Like Experiment 1, we examined misinformation recall probabilities for the highly confident responses and again found a protective effect of testing ($M$ = .08 for tested participants and $M$ = .26 for nontested participants), $t(37) = 2.41$, $p = .02$, $d = .78$. We also performed a conditional analysis examining final test accuracy depending on whether the participant successfully recalled an item on the initial test. Misinformation recall probability was very low ($M$ = .10) for the initially correct items but remained quite high for the initially incorrect items ($M$ = .31), $t(18) = 2.36$, $p = .03$, $d = .77$. We again examined the effects of encountering misinformation on retention of the initially correct items. Unlike the result in Experiment 1, presenting misinformation had no effect on final test correct recall probability here ($M$ = .93 for neutral and $M$ = .87 for misled), $t(19) = 1.39$, $p = .18$.

## General Discussion

In the present experiments, we sought to discover factors that govern whether retrieval exacerbates or reduces eyewitness suggestibility. We found that testing increased suggestibility when the misinformation was presented in a narrative; however, it protected against suggestibility when the misinformation was embedded in questions. These results help to clarify why, despite the apparently similar methodologies used in prior studies, testing has produced conflicting findings regarding its effects on eyewitness suggestibility.

More broadly, the present data provide an important empirical advancement to our knowledge regarding the complex interplay between retrieval practice and suggestibility. In particular, the conditions under which witnesses encounter misinformation can have a profound impact on their suggestibility. Along these lines, prior research has shown that people are far less susceptible to misinformation delivered by a peer than by an authority figure (Ceci, Ross, & Toglia, 1987; Dodd & Bradshaw, 1980; Lampinen & Smith, 1995; Smith & Ellsworth, 1987; Underwood & Pezdek, 1998), and that repeatedly encountering misinformation can increase its likelihood of being reported later (Foster, Huthwaite, Yesberg, Garry, & Loftus, 2012; Mitchell & Zaragoza, 1996; Zaragoza & Mitchell, 1996). As indicated in the Introduction, some research has also shown that encountering misinformation from a narrative or a question can influence its effectiveness at altering memory reports, but the results are somewhat mixed. In the present study, a quick comparison between data from the two experiments reveals that presenting misinformation in a narrative or a question had virtually no impact on eyewitness suggestibility when no initial test had taken place. However, far more interestingly, initial testing dramatically altered the relationship between misinformation presentation method and its effectiveness in inducing false recall. Whereas testing exacerbated suggestibility when misinformation was

presented in a narrative, it reduced suggestibility when misinformation was presented in a question. How can we account for these results?

We hypothesized that testing might reduce suggestibility if misinformation is presented in a question because attempting to answer a question may cause participants to retrieve other related information in memory, and the tested participants would be more likely (than the nontested subjects) to notice the inconsistency between the misinformation and what they remembered from the witnessed event. Consistent with this hypothesis, the testing effect in Experiment 2 was found only for items that were correctly recalled during the initial test. This finding suggests that conflict detection is driven primarily by people comparing their memory of what was recalled on the initial test (and not just what was seen in the video) with the misinformation. In the introduction, we mentioned that initial testing might enhance encoding of the misinformation and thus results in greater suggestibility (Gordon & Thomas, 2013). When misinformation is embedded within a narrative, initial testing enhances the attention participants give to the misleading, critical details. Although, the same may be true when misinformation is presented in misleading questions; misinformation presented in questions may be easier for tested participants to detect a conflict. Therefore, this test-enhanced encoding may work in different ways depending on the how misinformation is presented (i.e., enhancing encoding of the misinformation when it is presented in a narrative and conflict detection when it is embedded in questions). This remains a viable explanation for our findings.

In an attempt to keep our experimental design as consistent with former studies as possible, the misinformation was presented in a story-like audio narrative that recapped much of the witnessed event in Experiment 1. In contrast, the misleading questions were presented in a similar fashion to previous studies that demonstrated a testing effect (e.g., Pansky & Tenenboim,

2011). Here, the questions were presented visually, participants were given 30 sec to answer each question, and the 14 questions appeared without any background or contextual information about the story of the witnessed event. Due to these differences, we cannot ascertain that the retrieval requirement in the questions was the sole contributor to the reversal of the RES effect. We are currently conducting follow-up experiments to further examine the mechanisms that underlie this interesting reversal. Another potential limitation is that we are making comparisons between studies. That is, the participants in Experiment 1 could differ from those in Experiment 2 because data for the two experiments were not collected simultaneously. However, we believe this is unlikely to have caused the different effects of testing in Experiments 1 and 2 because data for both experiments were collected within a four week period. Moreover, all participants were from the same university, participating for course credit and the mean age of the participants did not differ between the experiments, $t < 1$, $p = .69$.

An intriguing finding in Experiment 2 was that the tested participants displayed poorer performance on the misleading questions (for which they needed to recall the noncritical details) than the nontested participants. One interpretation of this finding is that retrieving the critical details during the initial test might have impaired recall of the related, noncritical information (i.e., retrieval-induced forgetting). However, we cautioned against such an interpretation for three reasons. First, as described earlier, the impairment in recall performance for the tested participants relative to the nontested participants was only marginally significant. Second, our materials were highly coherent and interconnected in nature, and materials like these are highly resistant to retrieval-induced forgetting (Chan, 2009; Chan & LaPaglia, 2011; Migueles & Garcia-Bajos, 2007). Third, and perhaps most intriguingly, the decrement in performance for the tested participants might be the result of these participants noticing the inconsistencies between

their memory of the video and the misinformation. Specifically, this conflict detection might have drawn attention away from the task at hand – retrieving the non-critical detail that was requested by the question. Such diversion of attention would thus lower performance during the misleading question phase.

**Concluding Remarks**

Perhaps due to the popularity and excitement surrounding the idea that testing is a powerful retention enhancer (e.g., Rawson & Dunlosky, 2011; Roediger & Karpicke, 2006), there is a recent surge of interest in applying retrieval practice to reduce eyewitness suggestibility and to improve eyewitness memory performance in general. In line with this notion, testing has been shown in some studies to be an effective means to reduce suggestibility, and the common explanation here is that testing boosts retention of the witnessed event, which in turn allows witnesses to reject subsequently presented misinformation (e.g., Pansky & Tenenboim, 2010). More surprisingly, however, is that testing can sometimes produce the opposite effect – it can make witnesses far more susceptible to misinformation (e.g., Chan et al., 2009). The stark contrast of these findings raises the question, "how to account for this empirical puzzle?" In the present experiments, we found that different misinformation presentation method can be a powerful driving force behind whether retrieval enhances or reduces suggestibility. Although the exact mechanisms that underlie these disparate results remain to be elucidated, we believe the present findings represent an important step toward a fuller understanding of the relationship between retrieval and eyewitness suggestibility.

References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and mechanisms of forgetting. *Journal of Memory & Language, 49,* 415-445. doi: 10.1016/j.jml.2003.08.006

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

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. Journal of Experimental Psychology, 116, 38–49.

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61,* 153-170. doi: 10.1016/j.jml2009.04.004

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.doi: 10.1037/a0021204

Chan, J. C. K., & LaPaglia, J. A. (2011). The dark side of testing: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied, 17,* 418-432. doi: 10.1037/a0025147

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

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

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. Journal of Verbal Learning and Verbal Behavior, 19, 695–704.

Foster, J. L., Huthwaite, T., Yesberg, J. A., Garry, M., & Loftus, E. F. (2012). Repetition, not number of sources, increases both susceptibility to misinformation and confidence in the accuracy of eyewitnesses. *Acta Psychologia, 139,* 320-326. doi: 10.1016/j.actpsy.2011.12.004

Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a Self-Administered Cognitive Interview. *Applied Cognitive Psychology, in press.* doi: 10.1002/acp.2828

Gordon, L. T., & Thomas, A. K. (2013). Testing potentiates new learning in the misinformation paradigm. Manuscript submitted for publication.

Gobbo, C. (2000). Assessing the effects of misinformation on children's recall: How and when makes a difference. *Applied Cognitive Psychology, 14,* 163-182. doi: 10.1002/(SICI)1099-1720(200003/04)14:2<163::AID-ACP630>3.0.CO;2-H

Lane, S. M. (2006). Divided attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology, 20,* 199-212. doi: 10.1002/acp.1177

Lampinen, J. M., & Smith, V. L. (1995). The incredible (and sometimes incredulous) child witness: Child eyewitnesses' sensitivity to source credibility cues. Journal of Applied Psychology, 80(5), 621–627. doi:10.1037/0021-9010.80.5.621

LaPaglia, J. A., & Chan, J. C. K. (2012). Retrieval does not always enhance suggestibility: Testing can improve witness identification performance. *Law and Human Behavior.* doi: 10.1037/h0093931

Lee, Y., & Chen, K. (2012). Post-event information presented in a question form eliminates the misinformation effect. *British Journal of Psychology, in press*. doi: 10.111/j.2044-8295.2012.02109.x

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

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19-31. doi: 10.1037/0278-7393.4.1.19

Memon, A., Zaragoza, M., Clifford, B. R., & Kidd, L. (2010). Inoculation or antidote? the effects of cognitive interview timing on false memory for forcibly fabricated events. *Law and Human Behavior, 34,* 105-117. doi: 10.1007/s10979-008-9172-6

Migueles, M., & Garcia-Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology*, *21*, 1157–1172.

Mitchell, K. J., & Zaragoza, M. (1996). Repeated exposure to suggestion and false memory: The role of contextual variability. Journal of Memory and Language, 35, 246–260.

Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing. *Memory & Cognition, 39,* 155-170. doi: 10.3758/s13421-010-0005-8

Pezdek, K., & Roe, C. (1995). The effect of memory strength on suggestibility. *Journal of Experimental Child Psychology, 60,* 116-128. doi: 10.1006/jecp.1995.1034

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437-447. doi: 10.1016/j.jml.2009.01.004

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*, 283–302.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long term-retention. *Psychological Science, 17,* 249-255. doi: 10.111/j.1467-9280.2006.01693.x

Saunders, J. (2009). Memory impairment in the weapon focus effect. *Memory and Cognition, 37,* 326-335. doi: 10.3758/MC.37.3.326

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

Smith, V. L., & Ellsworth, P. C. (1987). The social psychology of eyewitness accuracy: Misleading questions and communicator expertise. Journal of Applied Psychology, 72(2), 294–300. doi:10.1037/0021-9010.72.2.294

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the build-up of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1392–1399.

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

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

Underwood, J., & Pezdek, K. (1998). Memory suggestibility as an example of the sleeper effect. Psychonomic Bulletin & Review, 5(3), 449–453. doi:10.3758/BF03208820

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498-505. doi: 10.3758/BF03192720

Weiser, B. (2012, July). New Jersey Court issues guidance for juries about reliability of eyewitnesses. *The New York Times*. Retrieved from www.nytimes.com

Wilford, M. M., Chan, J. C. K., & Tuhn, S. J. (2012). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. Manuscript in preparation.

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18,* 1140–1147. doi: 0.3758/s13423-011-0140-7

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 934-945. doi: 10.1037/0278-7393.20.4.934

Zaragoza, M. S., Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science, 7,* 294-300. doi: 10.1111/j.1467-9280.1996.tb00377.x

Table 1

*Mean Probabilities (and Standard Deviations) of Correct, Misinformation, Other and No Responses on the Initial Test in Experiments 1 and 2*

| | Correct | Misinformation | No Response | Other |
|---|---|---|---|---|
| Experiment 1 | .64 (.13) | .07 (.07) | .05 (.06) | .24 (.11) |
| Experiment 2 | .58 (.14) | .03 (.04) | .07 (.08) | .32 (.13) |

Table 2

*Mean Probabilities (and Standard Deviations) of No Response and Other Responses on the Final Test in Experiments 1 and 2*

**No Response**

| | Experiment 1 (Narrative) | | Experiment 2 (Questions) | |
|---|---|---|---|---|
| | No-Test | Test | No-Test | Test |
| Neutral | .04 (.06) | .05 (.08) | .03 (.08) | .06 (.09) |
| Misled | .01 (.03) | .01 (.04) | .04 (.08) | .04 (.10) |

**Other Responses**

| | Experiment 1 (Narrative) | | Experiment 2 (Questions) | |
|---|---|---|---|---|
| | No-Test | Test | No-Test | Test |
| Neutral | .31 (.17) | .29 (.18) | .42 (.22) | .35 (.20) |
| Misled | .24 (.14) | .17 (.13) | .24 (.20) | .26 (.22) |

Table 3

*Mean Probabilities (and Standard Deviations) of Correct, No Response, and Other Responses on the Misleading Questions Test in Experiment 2*

| | Correct | No Response | Other |
|---|---|---|---|
| No-Test | .82 (.07) | .04 (.05) | .14 (.09) |
| Test | .76 (.15) | .08 (.09) | .17 (.10) |

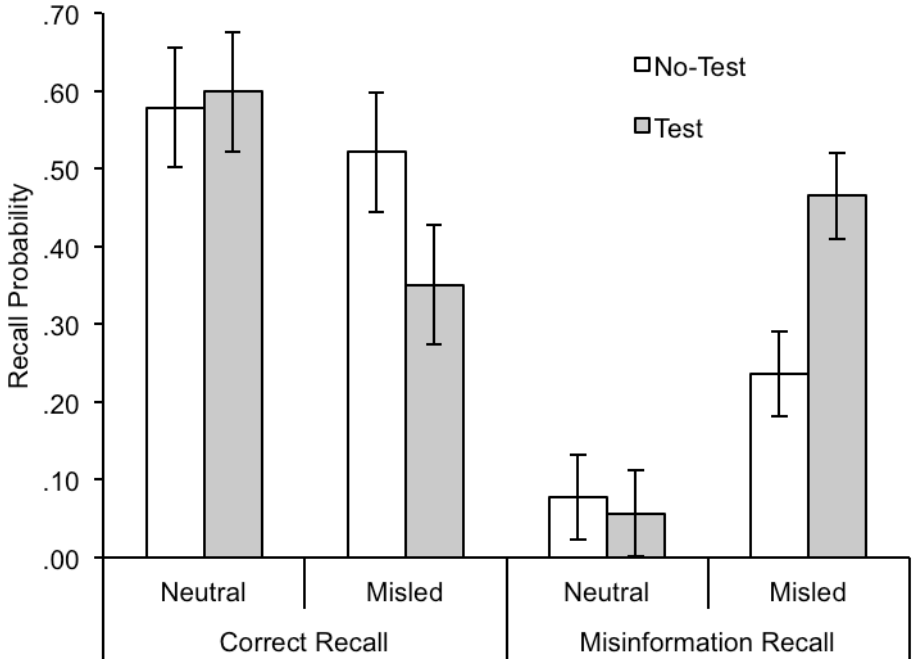


*Figure 1*. Experiment 1 final test correct and misinformation recall probabilities. Error bars indicate 95% confidence intervals.

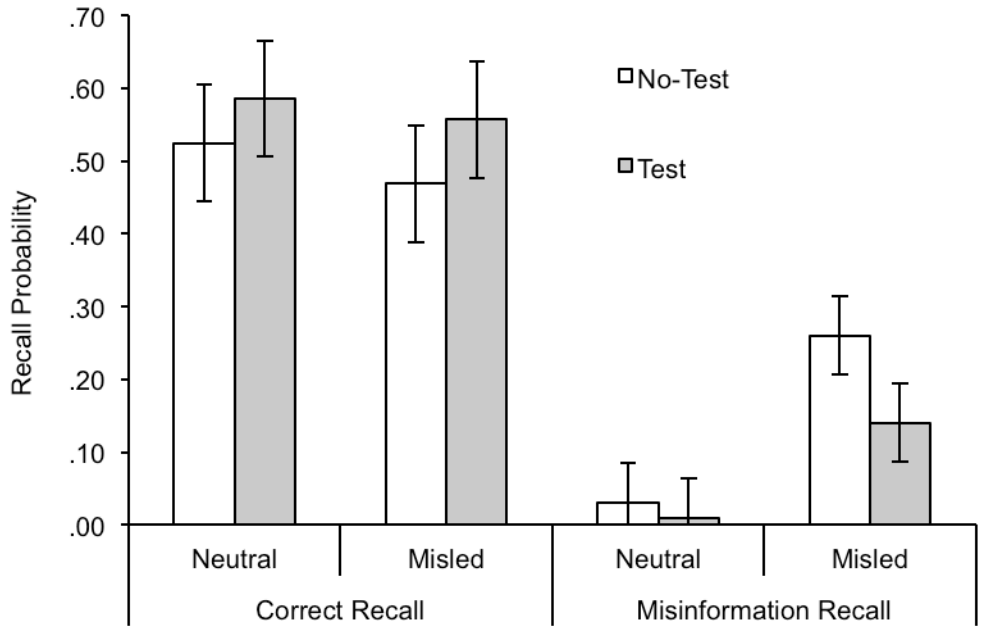


*Figure 2.* Experiment 2 final test correct and misinformation recall probabilities. Error bars indicate 95% confidence intervals.