

Retrieval Can Increase or Decrease Suggestibility Depending on How Memory is Tested: The
Importance of Source Complexity

Jason C. K. Chan

Miko M. Wilford

Katharine L. Hughes

Iowa State University

Address correspondence to:

Jason C. K. Chan
Department of Psychology
Iowa State University
Ames, IA 50011 USA

ckchan@iastate.edu
(515) 294-6946

Abstract

Taking an intervening test between learning episodes can enhance later source recollection. Paradoxically, testing can also increase people's susceptibility to the misinformation effect – an effect termed retrieval-enhanced suggestibility (RES, Chan, Thomas, & Bulevich, 2009). We conducted three experiments to examine this apparent contradiction. Experiment 1 extended the RES effect to a new set of materials. Experiments 2 and 3 showed that testing can produce opposite effects on memory suggestibility depending on the complexity of the source test. Specifically, retrieval facilitated source discriminations when the test contained only items with unique source origins. But when the source test included some items that had appeared in multiple sources, the intervening test actually increased source confusions. These results have implications for a wide variety of learning situations. We focused our discussion on eyewitness memory, source complexity, and reconsolidation.

Retrieval Can Increase or Decrease Suggestibility Depending on How Memory is Tested: The Importance of Source Complexity

Over 30 years ago, the seminal work by Loftus and her colleagues (e.g., Loftus & Palmer, 1974) highlighted the damaging power of misinformation on the accuracy of eyewitness memory reports. With today's round-the-clock news coverage, misinformation proliferates in the media. Therefore, investigations on how misinformation can affect people's memory are even timelier now than when the concept was originally introduced. In the present paper, we examined whether retrieval practice, or testing, can be used to reduce people's susceptibility to the misinformation effect.

According to the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993), manipulations that enhance source memory should reduce the influence of misinformation (Lindsay, 1990). Testing is one such manipulation. The testing effect is one of the most well documented empirical phenomena in cognitive psychology over the past decade (Rawson & Dunlosky, 2011), and its benefits extend well beyond strengthening memory of the retrieved items. In multi-list learning experiments, testing has repeatedly been found to augment later source retrieval. For example, testing can facilitate recollection of contextual information (Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010; Chan & McDermott, 2007; Verde, 2004) and can insulate against the buildup of proactive interference (Jang & Huber, 2008; Pastotter, Schicker, Niedernhuber, & Bauml, 2011; Robbins & Irvin, 1976; Szpunar, McDermott, & Roediger, 2008). Pastotter et al. proposed that testing promotes temporal (or list) segregation because performing retrieval between study cycles produces internal context change. If testing enhances source discrimination, then it should reduce eyewitness suggestibility. Paradoxically, some recent papers have reported the opposite. That is, taking a memory test of the witnessed

event actually increases people's suggestibility to subsequently presented misinformation – a finding termed retrieval-enhanced suggestibility or RES (e.g., Chan, Thomas, & Bulevich, 2009). Briefly, participants in these experiments first view an event and then either take an initial memory test (without any misleading suggestions) or perform some distractor task. After a retention interval, participants listen to an audio narrative that contains both correct and misleading information, and then they complete a final test. The critical finding is that initial retrieval increases recall and recognition of the misinformation. A possible explanation for RES is that initial testing potentiates learning of the misinformation (Wissman, Rawson, & Pyc, 2011), and people respond based on retrieval fluency (Benjamin, Bjork, & Schwartz, 1998) without spontaneously considering the source of retrieved information (Lindsay & Johnson, 1989). Because initial testing enhances new learning, it paradoxically increased accessibility of the misinformation.

Consistent with this retrieval fluency hypothesis, providing subjects with a (post-encoding) warning about the veracity of the misinformation narrative eliminates the RES effect (Thomas, Bulevich, & Chan, 2010). Indeed, the researchers noted that "... when provided with a warning, subjects took time to evaluate the source of the retrieved information, and this effortful process led to increased accuracy" (p. 156). Although this is a reasonable conclusion, one cannot be certain that the warning actually increased the likelihood of deliberate source monitoring because Thomas et al. never tested source memory. Furthermore, it is difficult to pinpoint the mechanisms underlying the effects of warning. Aside from more careful source monitoring, warning can reduce the misinformation effect by tightening participants' response criterion (Echterhoff, Groll, & Hirst, 2007). Alternatively, a warning can increase recognition of the original detail as long as participants can reject the misinformation, and this can occur even if

participants had failed to encode the original detail (for a thorough discussion, see McCloskey & Zaragoza, 1985). Moreover, participants may interpret a warning as permission to forget the narrative, consequently reducing accessibility of the misinformation (MacLeod, 1998). Because of the ambiguity inherent in a warning manipulation, the processes by which a warning eliminates the RES effect remain unknown. In the present research, we tested source memory directly.

The Present Experiments

Three experiments were conducted to examine the relation between testing and source monitoring in the context of eyewitness memory. Specifically, we propose that the effects of testing on subsequent witness suggestibility can be *positive* or *negative* depending on retrieval conditions. It is possible that testing would reduce suggestibility only when participants are required to make deliberate source judgments during retrieval, but testing would exacerbate suggestibility if participants instead respond based on fluency. In addition, we attempted to extend the generalizability of the RES phenomenon. To date, the same critical event was used in all existing investigations that reported RES (e.g., Chan et al., 2009; Thomas et al., 2010). Notably, the witnessed event in previous studies was a ~40 min video depicting a terror attack with multiple subplots (e.g., a family gathering, the rebellious behavior of a teenager, a presidential election, and a terrorist attack that involved a plane explosion). Arguably, such a complex event is not representative of the shorter events experienced by many eyewitnesses. In the current study, we used a much shorter (~8 min) video event. In addition, this video followed the actions of a single protagonist who stole a diamond from a heavily guarded museum. Indeed, other than the fact that they both included a crime, there was very little in common between the

video used in previous RES studies and the present one. This new set of materials helps to extend the generalizability of RES from a long, complex event to a short, simpler event.

All experiments had the same design. First, participants watched a video that depicted a museum burglary. They then either took an immediate cued recall test of the video (i.e., the *test condition*) or completed a distractor task (i.e., the *no-test condition*). Following a retention interval, all participants listened to a narrative containing primarily correct information in addition to six pieces of misinformation. After an additional retention interval, participants took the final test, which was cued recall for Experiment 1 and source recognition for Experiments 2 and 3.

Experiment 1

Method

Participants. Eighty undergraduate students from Iowa State University participated in this experiment for research credits, with 40 participants each in the test and no-test conditions.

Materials and Procedure. Participants first viewed the burglary video, which was an 8 min segment from the film “The Return of the Pink Panther.” They were told to pay close attention to the video but were not informed explicitly about any memory tests. After the video, participants in the test condition were administered an 18-question cued-recall test (see the Appendix for all the questions). No misinformation was presented during the initial test. Instead, the initial test served as retrieval practice and provided an assessment for the “immediate memory” of the event. Participants had 30 sec to type in an answer for each question. They were then asked to rate their confidence on a scale from 1 (not very confident) to 5 (very confident). Instead of the cued recall test, participants in the no-test condition played the videogame Tetris as a filler task.

All participants then watched a 10 min distractor video before listening to the post-event narrative via headphones. This narrative was ~6 min long and included 6 items of misinformation (all other information was accurate). All of the misinformation included alterations of details in the video (i.e., contradictory misinformation). For instance, one of the misinformation items involved the misrepresentation of a diamond stand that appeared in the video. Two versions of the narrative were created, which were identical with the exception of the six misinformation items (e.g., Version A: the diamond stand sits on a pedestal with *a pair of dolphins* sculpted into it; Version B: the diamond stand sits on a pedestal with *a number of cupid-like creatures* sculpted into it; Correct Answer: the diamond stand was made of glass with no carvings). The version of the narrative presented was counterbalanced across participants. In addition to the misinformation, the narrative included correct information regarding six details queried in the initial test (and the final test). These are referred to as *consistent items*.

Next, participants completed a working memory task as a distractor and then played the videogame Tetris for a total of 25 min. Afterwards, participants were presented with the final cued recall test, which was identical to the initial test. On this test, six questions queried details that had been presented inaccurately in the audio narrative (*misinformation items*), six queried details that appeared in the video and were also accurately described in the narrative (*consistent items*), and six queried details that appeared in the video but were not mentioned in the narrative (*control items*). To be clear, although participants in the testing group had received the same test earlier, the categorization of these items became meaningful only after the audio narrative. Participants were instructed specifically to answer all questions based on what they remembered from the video. Note that unlike Chan et al. (2009), we were unable to use the same detail for the different item types due to the brief nature of our video. Instead, the consistent, control, and

misinformation items were always about different details of the video (e.g., the detail regarding the diamond stand was always contradicted by misinformation in the narrative, see the Appendix); therefore, we did not compare performance across item types. Note that this is not of concern because the critical variable of interest here is testing, not item type.

Results and Discussion

Data from the recall tests were coded into separate categories for analyses. For all question types, the answers were classified as Correct, Other Intrusions (i.e., intrusions that did not match either the correct answer or the misinformation), and No Answer (i.e., questions that were left blank or responded with “I don’t know”). For the misinformation questions, there was an additional coding category – false recall (of the misinformation).

Data from the initial test are presented at the top of Table 1. The most important finding here is that the base rate false recall probability was, as expected, quite low (.09). There were also large differences in accurate recall probabilities between the consistent, control, and misinformation items, $F(2, 27) = 98.89, p < .01, \eta_p^2 = .88$, with the latter being the most difficult. Note again that this is not a problem because the comparison of interest here is between the testing conditions, not across the item types. We now present data from the final cued recall test, and we focus our analyses and discussions on correct recall and false recall (see Table 2 for means). All inferential analyses used an alpha level of .05. Effect size is reported using Cohen’s d . Note again that we did not compare recall probabilities across item types because they were about different details in the video. Instead, the critical comparisons were always between the no-test and test conditions. For **correct recall probabilities**, there was a testing effect for the control items (.71 for no-test and .82 for test), $t(78) = 2.61, d = .59, p = .01$, but there was a marginally significant reversed testing effect for the misinformation items (.25 for no-test and

.18 for test), $t(78) = 1.90$, $d = .43$, $p = .06$. In fact, initial testing nearly doubled the *false recall probability* (from .30 to .59), $t(78) = 5.14$, $d = 1.15$, $p < .01$. Clearly, we have extended the general pattern of RES to this new and very different set of materials.

An additional category of response deserves some attention. Specifically, researchers typically treat recall of misinformation as a “special” response due to its legal relevance, but from a practical perspective, it is often difficult to distinguish between misinformation-based errors and other idiosyncratic intrusions (i.e., in the courtroom, they are both errors). We investigated whether systematic differences exist between these two types of memory errors by examining their confidence. To ensure that an adequate number of responses were included in this analysis, we compared the False (.30) and Other (.38) responses in the no-test condition. Notably, false recall of misinformation was characterized by much higher confidence (4.06) than other intrusions (2.97), $t(31) > 3.66$, $d = .96$, $p < .01$. Remarkably, the confidence exhibited by false recall was even higher than accurate recall (3.57), $t(28) = 2.14$, $d = .47$, $p = .04$. Not only do these results suggest that there are systematic differences between the False and Other responses, they also highlight the highly problematic nature of misinformation from a legal perspective, because jurors tend to judge the accuracy of a statement by its perceived confidence (Leippe, Manion, & Romanczyk, 1992; Penrod & Cutler, 1995).

Experiment 2

Based on the retrieval fluency hypothesis, RES occurs because participants do not consider the source of a retrieved memory in the recall test. If this is the case, asking participants explicitly to provide source information during retrieval should reduce the RES effect. In fact, given the benefits that testing confers on source monitoring (Chan & McDermott, 2007) and proactive interference (Szpunar et al., 2008), a testing effect for source memory should emerge.

To provide a critical test of this idea, we forced participants to engage in deliberate source monitoring by giving them a source recognition final test.

Method

Participants. Eighty students participated in this experiment, with 40 in each test condition.

Materials and Procedure. All experimental protocols were identical to Experiment 1 except that the final test required source recognition instead of recall. The source test included 18 questions, with six each targeting the consistent, control, and misinformation items. For each test question, participants were provided with a question and an answer printed below.

Participants were asked to judge *the source of the answer* by selecting one of four options: a) video only, b) audio only, c) both video and audio, d) neither (i.e., the answer was new). The correct source judgments for the consistent, control, and misinformation items were “both,” “video only,” and “audio only,” respectively. Note that the source test is distractor-free, so the “neither” judgment was always incorrect.

Results and Discussion

As expected, the initial test data for Experiment 2 (see Table 1) were very similar to those for Experiment 1. Once again, base rate false recall probability was low (.04). We now present data for the final test.

We conducted planned comparisons for results of the source recognition final test (see Table 3). Overall, testing did not affect the ability for participants to correctly identify the source of the consistent and control items, $t_s < 1.17$, $p_s > .24$, $d < .27$. More important are the findings regarding the misinformation items. Here, testing did not affect the probability of correct source identification, $t < 1$, $d = .05$. Surprisingly, however, testing increased the likelihood for

participants to erroneously claim that a piece of misinformation was presented in *both* the video and the narrative – that is, testing increased source misattributions – an RES effect (No-test = .26, Test = .39), $t(78) = 2.43$, $p = .02$, $d = .55$. Therefore, despite the final test requiring subjects to make source discriminations, testing again increased the misinformation effect. This finding indicates that RES can occur even when retrieval conditions explicitly discourage responding based on fluency; instead, it indicates that participants actually possessed illusory recollection of the misinformation being part of the video. Further substantiating this claim, participants did not choose the “both” option simply when they were uncertain. Both the tested and nontested participants were more confident when they chose the incorrect “both” option ($M = 4.20$) than when they chose the correct “audio only” option ($M = 3.72$), $t(30) = 2.42$, $p = .02$, $d = .56$.

Intriguingly, the RES effect on source confusion is unidirectional, such that testing increased only “both” judgments for the misinformation items but not for the control items, $t < 1$, $d = .17$. These findings are consistent with those reported by Hupbach, Gomez, and Nadel (2009, 2011), who suggest that retrieval triggers a memory updating mechanism that allows new information to better bind with memory of the recently-retrieved event (see also Brown, Brown, Mosbacher, & Dryden, 2006). The idea is that a formerly stabilized memory becomes labile upon reactivation and must undergo a process of reconsolidation. During reconsolidation, new information is more easily integrated into the original memory representation (for recent reviews on the topic, see Hardt, Einarsson, & Nader, 2010; Nader & Einarsson, 2010). In the present context, initial retrieval may cause the subsequently learned misinformation to be better integrated into memory of the video event, thereby increasing “both” judgments. We consider this possibility more thoroughly following presentation of the results from Experiment 3.

Experiment 3

To further understand the influence of prior retrieval on subsequent source memory, we first consider an empirical puzzle: Why are the benefits of testing on source retrieval so elusive in eyewitness memory studies (for an exception, see Pansky & Tenenboim, 2011) but readily obtained in word list studies? We suspect that this discrepancy can be attributed to a major difference in the materials used in these two lines of research. In list learning studies, information in the two sources (e.g., word lists 1 and 2) virtually never overlaps (for an exception, see Dodhia & Metcalfe, 1999), such that the two learning sources contain unique items (i.e., items studied in list 1 never appear in list 2 and vice versa, e.g., Bodner & Lindsay, 2003; Dobbins & McCarthy, 2008; Dodson, Holland, & Shimamura, 1998). However, in misinformation studies, items in the two sources almost always contain some degree of overlap, because the misinformation is presented in the context of the original event (Lane, Roussel, Villa, & Morita, 2007; Lindsay & Johnson, 1989; Zaragoza & Lane, 1998).

This difference in methodology is important because source attributions are much more challenging when information in the two sources overlap. In list learning studies, participants are often asked to judge whether an item was presented in the first or second source (i.e., list 1 or list 2). Conversely, in misinformation studies, because of the item overlap, participants must grapple with the possibility that an item could have appeared in *both* sources. When memory for sources with unique items is tested, remembering the item from *one* source provides diagnostic information (e.g., remembering the item was learned in list 2 means that it was not learned in list 1). However, when memory for sources containing some identical items is tested, remembering that an item definitely appeared in one source (e.g., the narrative) does not provide any diagnostic information about whether that item also appeared in the other source (e.g., the video). When participants encounter the misinformation, they can reject the “both” option only if they a)

remember both the items presented in the video and narrative (e.g., “I remember that the video showed a glass diamond stand and the narrative described a marble stand.”), b) remember that the misinformation was presented in the narrative and something else was presented in the video (e.g., “I remember that the narrative mentioned a marble stand, but the video showed something different.” Tousignant, Hall, & Loftus, 1986), or c) remember the item presented in the video and that it did not match the misinformation (e.g., “I remember the video showed a glass stand, not a marble stand”).

In Experiment 3, we examined the hypothesis that the testing effect on source retrieval is tied, at least partially, to the non-overlapping source test format. Accordingly, participants were simply asked to indicate whether an item was presented in the video or audio only. We reasoned that because initial testing reduced proactive interference, it should strengthen the ability for participants to remember that the misinformation was presented during the narrative. However, as described previously, this is of limited use when one is confronted with the “both” possibility – just because one can clearly remember an item from the narrative does not mean that the item did not also appear in the video. Consequently, removal of the “both” option should allow the testing effect on source monitoring to emerge.

Method

Participants. Forty students participated in this experiment, with 20 in each test condition.

Materials and Procedure. All experimental protocols were the same as Experiment 2 with the following exception: To make the test items commensurate with the source options available (i.e., video only vs. audio only), the final test included only questions for the control

items and the misinformation items. Therefore, the source test contained 12 rather than 18 questions.

Results and Discussion

The initial recall test data are presented at the bottom of Table 1. Again, base rate false recall probability was low (.08). We now report data from the source recognition final test. Because this test involved binary choices, only one contrast was performed for each item type (see Table 4). Similar to Experiment 2, initial testing did not affect correct source attributions of the control items, $t(38) = 1.04$, $p = .31$, $d = .33$. More importantly, unlike Experiment 2, a testing effect on correct source judgments emerged for the misinformation items (No-test = .64, Test = .80), $t(38) = 2.57$, $p = .02$, $d = .81$. That is, initial testing increased the likelihood for participants to correctly claim that a piece of misinformation was presented in the narrative. Because Experiments 2 and 3 were identical until final retrieval, the opposite effects of testing on source performance between these experiments cannot be attributed to a difference in the encoding or retention of source information, but only to a difference in the decision processes involved in source judgments. Specifically, these data suggest that because testing reduced proactive interference during “misinformation learning,” participants could more easily remember that the misinformation was definitely encountered during the narrative. When participants no longer had to distinguish between the “audio only” and “both” options, the correct answer became apparent.

How do these results reconcile with those from Experiment 2, which suggests that participants possessed illusory recollections of the misinformation from the video? One explanation is that initial testing increased the likelihood that one would remember the misinformation in the video (i.e., testing facilitated integration of the misinformation into the

memory of the video event during reconsolidation), but testing also increased participants' ability to remember that the misinformation was presented during the audio narrative (because testing reduced proactive interference). Although these two explanations may seem incompatible, they are not mutually exclusive; because one can remember *two presentations* of the misinformation – one from the video and one from the narrative (which corresponds to the “both” option in Experiment 2) – and such remembrances can be affected by different processes.

Taken together, our results show that the testing effect on source memory is not a universal phenomenon (see also Brewer et al., 2010), and even more importantly, they highlight that variations in the complexity of the source test can produce empirical outcomes that suggest dramatically different, even opposite, conclusions. This complexity component is oft-overlooked in research on source monitoring (for notable exceptions, see Dobbins & McCarthy, 2008; Dodson & Johnson, 1993; R. L. Marsh & Hicks, 1998; Parks, Murray, Elfman, & Yonelinas, 2011; Qin, Raye, Johnson, & Mitchell, 2001).

General Discussion

In three experiments, we first showed that the RES effect is not tied to the particular set of materials used in previous research. More importantly, we showed that testing can increase or decrease source misattributions depending on the way the test is constructed. These results underscore the importance of retrieval conditions in revealing the effects of testing on subsequent memory. We now discuss their theoretical implications.

Why did testing increase “both” attributions? An intriguing possibility is that prior retrieval may facilitate assimilation of the new (mis)information into memory of the original event, thus increasing selection of the “both” option (Hupbach et al., 2009, 2011). In Hupbach et al.'s experiments, participants encountered two sets of objects during separate experimental

sessions. If participants were reminded of the first session before they encountered the new objects, they were more likely to later misattribute these objects to Session 1. An important distinction between the Hupbach et al. experiments and ours was that their participants did not reactivate memory for the original episode with a memory test. Instead, participants were reminded of the original experimental session without explicit retrieval.¹ Therefore, no testing effect was expected for their experiment. Nonetheless, the reconsolidation theory provides a viable explanation for our data.²

It is surprising that RES remained robust in a source monitoring test, whereas it was eliminated altogether with a warning (Thomas et al., 2010). This discrepancy suggests that warning did not reduce RES by prompting careful source monitoring during retrieval. Instead, it reduced RES by biasing participants against reporting items from the post-event narrative. Because testing enhances participants' ability to remember that an item was presented in the narrative (as indicated by the results from Experiment 3), it should therefore increase their ability to selectively withhold the misinformation when they are given a warning.

A testing effect on source memory was revealed in the non-overlapping source test in Experiment 3. One explanation for this benefit is that taking an intervening test promotes list isolation (Jang & Huber, 2008), which facilitates subsequent source identification performance. If this is the case, then why is testing not capable of enhancing source judgments in the overlapping source test in Experiment 2? Although it is possible that list isolation is ineffective when the same items had appeared across two learning episodes, it is still unclear why list isolation failed to facilitate source identification of the misinformation in Experiment 2, given that the misinformation appeared in only the narrative. An alternative, more exciting possibility is that whether testing triggers event isolation or assimilation depends on whether participants

perceive the two learning sources as related. If participants believe that the two learning episodes are about the same event, prior retrieval would promote memory updating (i.e., assimilation); however, if participants perceive the two learning episodes as unrelated, prior retrieval would lead to temporal segregation. This hypothesis highlights the complex interplay between retrieval and subsequent encoding, and it points to a mechanism that adapts to perceived task demands. We believe this hypothesis has substantial explanatory power and can help resolve the empirical puzzle of whether testing enhances or impairs subsequent source recollection. Clearly, further research is needed to verify this tentative hypothesis.³ To be clear, we believe that our procedure produced list assimilation (as indicated by the results of Experiment 2), even though the results of Experiment 3 gave the apparent impression of list isolation.

Although the current paper is framed under the umbrella of eyewitness memory, we believe these data have implications for other learning domains as well. For example, students are sometimes exposed to misinformation in the classroom (e.g., from another student or from foils in a multiple choice exam) and in movies (see e.g., E. J. Marsh & Fazio, 2006). Moreover, people are often exposed to misleading information in the media, especially during political election seasons, when politicians misrepresent their rivals in debates and advertisements. In fact, political misrepresentations are so ubiquitous that they prompted the creation of multiple political fact checkers (e.g., Fackcheck.org, Politifact.com).

In 2010, the Supreme Court ruled that the First Amendment restricts government from prohibiting political advertisements funded by corporations (*Citizens United v. Federal Election Commission*). In its explanation, the Supreme Court claimed that proper disclosure of funding agencies enables people to judge the accuracy and importance of a message based on its source

(Liptak, 2011). Troublingly, the present data and other research on source monitoring have clearly shown that people frequently forget the source of retrieved information (for related phenomena, see the illusory truth effect, Begg, Anas, & Farinacci, 1992; Henkel & Mattson, 2011; and the sleeper effect, Kumkale & Albarracin, 2004). Given its relevance to real-world learning phenomena, it is imperative that research continues to investigate the mechanisms that exacerbate and protect against source confusions.

References

- Begg, I., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*, 446-458.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55-68.
- Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, *48*, 563-580.
- Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory*, *18*, 385-393.
- Brown, A. S., Brown, C. M., Mosbacher, J. L., & Dryden, W. E. (2006). Proactive and retroactive effects of negative suggestion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1234-1243.
- 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., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 431-437.
- 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.

- Dobbins, I. G., & McCarthy, D. (2008). Cue-framing effects in source remembering: A memory misattribution model. *Memory & Cognition*, *36*, 104-118.
- Dodhia, R. M., & Metcalfe, J. (1999). False memories and source monitoring. *Cognitive Neuropsychology*, *16*, 489-508.
- Dodson, C. S., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1121-1136.
- Dodson, C. S., & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology*, *106*, 541-557.
- Echterhoff, G., Groll, S., & Hirst, W. (2007). Tainted truth: Overcorrection for misinformation influence on eyewitness memory. *Social Cognition*, *25*, 367-409.
- Hardt, O., Einarsson, E. O., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology*, *61*, 141-167.
- Henkel, L. A., & Mattson, M. E. (2011). Reading is believing: The truth effect and source credibility. *Consciousness and Cognition*, *20*, 1705-1721.
- Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: Updating or source confusion? *Memory*, *17*, 502-510.
- Hupbach, A., Gomez, R., & Nadel, L. (2011). Episodic memory updating: The role of context familiarity. *Psychonomic Bulletin & Review*, *18*, 787-797.
- Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 112-127.

- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.
- Kumkale, C. T., & Albarracin, D. (2004). The sleeper effect in persuasion: A meta-analytic review. *Psychological Bulletin, 130*, 143-172.
- Lane, S. M., Roussel, C. C., Villa, D., & Morita, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source-monitoring errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 1131-1142.
- Leippe, M. R., Manion, A. P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and childrens' memory reports? *Journal of Personality and Social Psychology, 63*, 181-197.
- Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1077-1083.
- Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17*, 349-358.
- Liptak, A. (2011, September 19). A blockbuster case yields an unexpected result, *The New York Times*, p. A13.
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585-589.
- MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1-57). Mahwah, NJ: Erlbaum.

- Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, *34*, 1140-1149.
- Marsh, R. L., & Hicks, J. L. (1998). Test formats change source monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1137-1151.
- McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*, 1-16.
- Nader, K., & Einarsson, E. O. (2010). Memory reconsolidation: an update. *Annals of The New York Academy of Sciences*, *1191*, 27-41.
- Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing. *Memory & Cognition*, *39*, 155-170.
- Parks, C., Murray, L. J., Elfman, K., & Yonelinas, A. P. (2011). Variations in recollection: The effects of complexity on source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 861-873.
- Pastotter, B., Schicker, S., Niedernhuber, J., & Bauml, K.-H. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 287-297.
- Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: assessing their forensic relation. *Psychology, Public Policy and Law*, *1*, 817-845.
- Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1110-1115.

- 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.
- Robbins, D., & Irvin, J. R. (1976). The priority effect: Test effects on negative transfer and control lists. *Bulletin of the Psychonomic Society*, *8*, 167-168.
- 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.
- 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.
- Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition*, *14*, 329-338.
- Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition*, *32*, 1265-1272.
- 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.
- Zaragoza, M., & Lane, S. M. (1998). Processing resources and eyewitness suggestibility. *Legal and Criminological Psychology*, *3*, 305-320.

Footnotes

¹ To avoid the testing effect from influencing their results, Hupbach et al. (2009) actually prohibited their participants from recalling specific objects encountered during the original experimental session.

² Although Chan and Langley (2011) reported that interference during reconsolidation is not necessary for RES to occur, their data do not rule out the possibility that reconsolidation can play a role under some circumstances (e.g., when the delay between initial retrieval and misinformation learning is short, as in the present experiments).

³ One may wonder how this hypothesis would reconcile with Hupbach et al.'s findings, where initial testing increased source confusions for two ostensibly unrelated sets of objects. Note again that subjects in their experiments never performed retrieval practice, so no temporal segregation can be expected.

Table 1.

Recall Probabilities During the Initial Test as a Function of Item Type, Response Type, and Experiment

	Response Type			
	Correct	Other Intrusions	No Answer	Base rate False Recall
Experiment 1				
Consistent Items	.85 (.17)	.13 (.16)	.02 (.05)	
Control Items	.75 (.24)	.16 (.23)	.10 (.11)	
Misinformation Items	.40 (.19)	.42 (.17)	.09 (.64)	.09 (.15)
Experiment 2				
Consistent Items	.81 (.19)	.16 (.17)	.03 (.06)	
Control Items	.76 (.20)	.19 (.15)	.05 (.13)	
Misinformation Items	.44 (.18)	.45 (.18)	.06 (.11)	.04 (.09)
Experiment 3				
Consistent Items	.78 (.21)	.18 (.15)	.04 (.09)	
Control Items	.68 (.19)	.19 (.19)	.13 (.12)	
Misinformation Items	.38 (.19)	.46 (.19)	.09 (.11)	.08 (.10)

Note. Standard Deviations Appear in Parentheses.

Table 2.

Recall Probabilities During the Final Test in Experiment 1 as a Function of Item Type and Response Type

	Response Type			
	Correct	Other Intrusions	No Answer	False Recall
Consistent Items				
No-test	.83 (.18)	.12 (.15)	.05 (.09)	
Test	.88 (.18)	.13 (.18)	.00 (.00)	
Control Items				
No-test	.71 (.18)	.22 (.17)	.07 (.08)	
Test	.82 (.19)	.11 (.17)	.07 (.09)	
Misinformation Items				
No-test	.25 (.16)	.38 (.17)	.07 (.09)	.30 (.21)
Test	.18 (.18)	.20 (.17)	.03 (.09)	.59 (.29)

Note. Standard Deviations Appear in Parentheses.

Table 3.*Proportions of Source Attributions in Experiment 2 as a Function of Item Type*

	Source Attribution			
	Video Only	Audio Only	Both	Neither
Consistent Items				
No-test	.10 (.14)	.03 (.06)	.84 (.16)	.03 (.08)
Test	.16 (.15)	.02 (.07)	.80 (.16)	.02 (.06)
Control Items				
No-test	.49 (.24)	.07 (.11)	.32 (.21)	.13 (.13)
Test	.53 (.28)	.05 (.09)	.36 (.21)	.07 (.14)
Misinformation Items				
No-test	.15 (.17)	.40 (.25)	.26 (.23)	.19 (.20)
Test	.12 (.15)	.42 (.26)	.39 (.26)	.07 (.15)

Note. Standard Deviations Appear in Parentheses. Correct Attributions are Shown in Bold and Italics Font.

Table 4.*Proportions of Source Attributions in Experiment 3 as a Function of Item Type*

	Source Attribution	
	Video Only	Audio Only
Control Items		
No-test	.86 (.16)	.14 (.16)
Test	.79 (.23)	.21 (.23)
Misinformation Items		
No-test	.36 (.17)	.64 (.17)
Test	.20 (.21)	.80 (.21)

Note. Standard Deviations Appear in Parentheses. Correct Attributions are Shown in Bold and Italics Font.

Appendix

Misinformation Items

1. As the burglar enters the roof of the museum, the entrance is of a particular shape. What shape was it?

[Correct Answer: Octagonal, Misinformation Version A: Circular, Misinformation Version B: Square]

2. Upon entering the museum, the burglar finds himself in a dimly lit hallway area, what color were the walls painted?

[Correct Answer: White or Off-White, Misinformation Version A: Beige, Misinformation Version B: Grey]

3. Describe the pedestal, which is used to hold the large baseball-shaped diamond in the big room.

[Correct Answer: Glass, hourglass-shaped, Misinformation Version A: Marble, Misinformation Version B: Marble cupids dolphins]

4. A piece of artwork was present on the ceiling of the big room (where the burglar stole the diamond). Describe what the artwork looked like and be as specific as possible (e.g., what patterns were present in the artwork, etc.).

[Correct Answer: Wooden, brown circular patterns, Misinformation Version A: Knights charging, Misinformation Version B: David & Goliath]

5. The glove had the letter P embellished on it, what color was the letter embellished on the glove?

[Correct Answer: Silver, Misinformation Version A: White, Misinformation Version B: Gold]

6. All the museum guards wear hats as part of their uniform, what color were the hats the guards wore?

[Correct Answer: Dark green, Misinformation Version A: Tan, Misinformation Version B: Dark brown]

Control Items

7. Two museum guards talk while the burglar is exploring the museum. They are standing in front of what appears to be a stained glass window. Please describe the design of this window as specifically as possible (e.g., colors, patterns, etc).

[Correct Answer: Simple shapes, yellow, red, blue]

8. In addition to the stained glass window, what large objects are near the two guards as they are talking?

[Correct Answer: Two Greek statues, white]

9. The dome-shaped glass case that covered the diamond had a handle on it, what shape is the handle of this case?

[Correct Answer: Moon-shaped, crescent]

10. A carpet was placed underneath the diamond stand in the large room and surrounded by velvet ropes. Describe the carpet (size, color, pattern, etc.).

[Correct Answer: Red, octagonal]

11. The alarm is eventually set off, who sets the alarm off and how?

[Correct Answer: The guard accidentally sets it off upon seeing the glove]

12. As the burglar escapes from the museum, the guards begin to shoot at him, how many guards were shooting at the burglar as he got away?

[Correct Answer: 3]

Consistent Items

13. The burglar begins on the roof of a building. How does the burglar get from the roof of the building he starts on, to the roof of the museum?

[Correct Answer: Slid down a cable with use of crossbow]

14. How does the burglar get under the alarm beams undetected when advancing to the room with the diamond?

[Correct Answer: Uses the crossbow to shoot a rope and pulls himself across]

15. After getting to the room with the diamond, how does the burglar remove the dome-shaped glass case covering the diamond?

[Correct Answer: Uses the crossbow to shoot a rope and make a pulley]

16. When the burglar is working to retrieve the diamond, what apparatus, if any, is used to retrieve the diamond?

[Correct Answer: Two silver, metal arms (2 & 3 pronged)]

17. Once the burglar steals the diamond, a glove is put in its place, what was the color of the glove that the burglar put in place of the diamond?

[Correct Answer: White]

18. How does one of the guards finally discover that there is something going wrong in the museum?

[Correct Answer: Sees the glove in place of the diamond]