Original Article

QJEP

Quarterly Journal of Experimental Psychology
2020, Vol. 73(3) 425–441

DOI: 10.1177/1747021819881948
qjep.sagepub.com
SAGE

Protective effects of testing across misinformation formats in the household scene paradigm

Rosemary S Pereverseff[1], Glen E Bodner[2] and Mark J Huff[3]

Abstract

Many studies have demonstrated retrieval-enhanced suggestibility (RES), in which taking an initial recall test after witnessing an event increases suggestibility to subsequent misinformation introduced via a narrative. Recently, however, initial testing has been found to have a protective effect against misinformation introduced via cued-recall questions. We examined whether misinformation format (narrative vs. cued-recall questions) yields a similar dissociation in a paradigm that, to date, has consistently yielded a protective effect of testing (PET). After studying photos of household scenes (e.g., kitchen), some participants took an initial recall test. After a 48-hr delay, items not presented in the scenes (e.g., knives/plates) were suggested either via narrative or questions. Regardless of the misinformation format, we found a PET on both initial-test-conditionalised free recall and source-monitoring tests. However, initial testing also yielded memory costs, such that suggested items reported on the initial test were likely to persist on a final recall test. Thus, initial testing can protect against suggestibility, but can also precipitate memory errors when intrusions emerge on an initial test.

Keywords

Misinformation effect; initial testing; retrieval-enhanced suggestibility; protective effect of testing; recall; source monitoring

Received: 16 November 2018; revised: 16 May 2019; accepted: 5 September 2019

Understanding the factors that influence memory accuracy is an important issue in both basic and applied research. The malleability of memory is highlighted by the misinformation effect—which occurs when individuals report falsely suggested details about a previous event (Loftus, Miller, & Burns, 1978; see Zaragoza, Belli, & Payment, 2007 for review). The reduction in memory accuracy following exposure to misinformation is of particular concern in eyewitness contexts. Exposure to misinformation can have severe consequences when eyewitnesses are later interviewed, or testify about their memory in court. Our study contributed to this issue by examining whether completing an initial memory test after an event increases or decreases susceptibility to misinformation, and whether the method through which misinformation is introduced modulates susceptibility.

In the standard misinformation paradigm (e.g., Loftus et al., 1978), participants are exposed to misleading information about a previously experienced event. On a later test, misleading details are reported or endorsed more often than when they were not suggested. The misinformation effect occurs whether misinformation about the event is embedded in a series of questions (e.g., Saunders & Jess, 2010) or in a narrative account (e.g., Takarangi, Parker, & Garry, 2006). The effect is robust and persists even when participants are provided with explicit warnings about misleading details (Chambers & Zaragoza, 2001; Eakin, Schreiber, & Sergeant-Marshall, 2003; Echterhoff, Hirst, & Hussy, 2005; Zaragoza, Lane, Ackil, & Chambers, 1997) or are required to specify the source of their retrievals at test (Huff, Davis, & Meade, 2013; Lindsay & Johnson, 1989; Mitchell, Johnson, & Mather, 2003).

[1]Department of Psychology, University of Calgary, Calgary, AB, Canada
[2]College of Education, Psychology and Social Work, Flinders University, Adelaide, SA, Australia
[3]School of Psychology, The University of Southern Mississippi, Hattiesburg, MS, USA

Corresponding author:
Mark J Huff, School of Psychology, The University of Southern Mississippi, Hattiesburg, MS 39406, USA.
Email: mark.huff@usm.edu

Given the deleterious effects of misinformation on memory and their potential ramifications, it is important to identify conditions that can help to inoculate memory. Methods such as warnings may not be feasible in eyewitness contexts where it is unknown whether eyewitnesses were exposed to misinformation, and may even discourage witnesses from reporting details for fear that they may be incorrect. Despite these limitations, research has identified some means of enhancing memory for a witnessed event. For example, shortening the delay between the original event and misinformation exposure can reduce the misinformation effect (Belli, Windschitl, McCarthy, & Winfrey, 1992), as can ensuring full attention is available during original encoding (Lane, 2006). Again however, most eyewitness situations do not offer control over such parameters.

More generally, what is needed are methods of reducing memory suggestibility that can be applied *after* an event, but *before* the witness is exposed to misinformation. Initial memory testing following an event is a potential candidate. Taking a memory test has been shown to be a powerful memory modifier that can enhance retention relative to additional study opportunities (Roediger & Karpicke, 2006). This benefit of testing is termed the retrieval-practice effect, and it has been ascribed to processes including enhanced memory organisation (Congleton & Rajaram, 2012) and increased generation of implicit "mediator" cues that can be used at retrieval (Carpenter, 2011; Pyc & Rawson, 2010), which in turn can slow forgetting rates (Kornell, Bjork, & Garcia, 2011; for reviews see Rawson & Dunlosky, 2011; Rowland, 2014).

Surprisingly, given the established benefits of retrieval practice, initial testing has been shown to *increase* suggestibility to misinformation—a pattern termed *retrieval-enhanced suggestibility* (RES; Thomas, Bulevich, & Chan, 2010; for a review see Chan, Manley, & Lang, 2017). The RES pattern was first reported by Chan, Thomas, and Bulevich (2009). After viewing an episode of the television series *24*, some participants completed an initial cued-recall test for details of the episode. Relative to a no-test condition, initial testing increased suggestibility to contradictory details about the episode that were later presented via an auditory narrative. Chan and Langley (2011) replicated this RES pattern when the final test occurred after a week delay. The RES pattern occurs with different types of video events (Chan, Wilford, & Hughes, 2012), with both longer (40 min) or shorter (8 min) durations (Chan et al., 2009; Wilford, Chan, & Tuhn, 2014), and when an initial cognitive interview (a method used to probe eyewitnesses) is used in place of an initial cued-recall test (LaPaglia, Wilford, Rivard, Chan, & Fisher, 2014). Counterintuitively, this RES pattern suggests that giving eyewitnesses an initial memory test may backfire and increase susceptibility to misinformation.

Two mechanisms have commonly been proposed to account for the RES pattern. The test-potentiated learning account posits that initial testing potentiates learning of subsequent information—including misinformation (Cho, Neely, Crocco, & Vitrano, 2017; Pastötter & Bäuml, 2014). Compared with a no-test control, initial testing is said to facilitate the encoding of misleading details, enhancing the misinformation effect on a final test. Consistent with this account, some studies showing RES have also shown that initial testing improves memory for correct event details presented with post-event information (i.e., *consistent items*; Chan & Langley, 2011; Chan & LaPaglia, 2011; Chan et al., 2017; Thomas et al., 2010, but see Chan et al., 2009). Importantly, however, these testing benefits were not global: event details that do not appear in post-event information (i.e., *control items*) did not show a retrieval-practice effect.

By a second account, initial testing increases endorsement of misinformation because participants consider discrepant misinformation details to be corrective feedback, leading them to pay greater attention to those details during the misinformation phase. Consistent with this possibility, Gordon and Thomas (2014) reported that RES was associated with longer reading times for statements containing misleading details (vs. statements without misleading details) in the narrative. Initial testing may have increased attention allocated to discrepant items during the misinformation phase, thus enhancing their encoding. These test-potentiated learning and attentional accounts are not mutually exclusive, and both mechanisms may contribute to RES.

Strikingly, another set of studies has yielded the opposite result, namely, that initial testing can *decrease* one's susceptibility to misinformation. A *protective effect of testing* (PET; Huff et al., 2013; Huff, Weinsheimer, & Bodner, 2016) was first reported by Loftus (1977). In her study, after viewing a car involved in an accident, some participants were asked to indicate the car's colour prior to receiving a misleading suggestion of a different colour. On a final test, these participants were more accurate at reporting the original colour relative to participants who did not take the initial colour test. More recently, Memon, Zaragoza, Clifford, and Kidd (2010) found that completing a guided cognitive interview about one's memory for an event reduced later suggestibility relative to completing the interview after exposure to misinformation. These and other studies (e.g., Gabbert, Hope, Fisher, & Jamieson, 2012; Lane, Mather, Villa, & Morita, 2001; Pansky & Tenenboim, 2011) have found a clear benefit of initial testing, in line with the literature showing retrieval-practice effects.

Our study was based on a *household scene paradigm* developed by Huff et al. (2013) that, to date, has also consistently yielded a PET pattern. Their participants studied a set of six images of household scenes (e.g., kitchen, bathroom), each of which contained many scene items. Some participants then completed an initial free recall task for the items from each scene. All participants were then exposed to a set of fake recall tests ostensibly completed by other participants, some of which introduced false suggested items that were not actually shown in the scenes. Participants then completed a final free recall test and a source-monitoring test. A significant retrieval-practice effect on the final recall test was found such that correct recall was enhanced by initial testing, but initial testing did not affect recall of misinformation. However, a PET pattern was found on the source-monitoring test, such that taking an initial test (vs. no test) resulted in fewer suggested items being attributed to the original photos. This pattern was replicated and extended by Huff et al. (2016), whose participants completed one or two initial recall tests and were exposed to misinformation either immediately or following a 48-hr delay. The PET pattern on the source-monitoring test replicated, and importantly, a PET pattern emerged on the free recall test after a delay. Taking two initial tests was not more effective than taking one. In sum, accumulating evidence shows that in some paradigms or some conditions, initial testing can reduce misinformation susceptibility.

Given that initial testing can either increase or decrease misinformation susceptibility, research has begun to focus on identifying the factors that determine whether a RES or PET pattern occurs. Thomas et al. (2010) warned participants that they may have been exposed to erroneous details prior to completing the final test. This warning eliminated RES, but it did not produce a PET pattern. Participants' ability to avoid reporting misinformation after a warning indicates that their memories for the correct event details were still intact (e.g., McCloskey & Zaragoza, 1985) rather than having been overwritten (Loftus et al., 1978). Thus, warning individuals about exposure to false items appears to influence whether a RES or no effect of initial testing occurs, rather than whether a RES or PET occurs.

A second potentially relevant factor is the type of initial/final memory test. The RES pattern has typically been obtained when the initial/final test was cued recall (e.g., Chan et al., 2017; Chan et al., 2009; Gordon, Thomas, & Bulevich, 2015; Thomas et al., 2010), though it has also been obtained when the initial/final test was free recall (Wilford et al., 2014); when a source-monitoring test was used (Chan et al., 2012); and when the initial test included the cognitive interview (LaPaglia et al., 2014). The PET pattern, too, has been found using a variety of initial/final test types including cued recall (Pansky & Tenenboim, 2011), free recall (Gabbert et al., 2012; Huff et al., 2016; Loftus, 1979), recognition (Loftus, 1977), and when participants are required to specify the source of their retrievals through a source-monitoring test (Huff et al., 2013; Huff et al., 2016; Johnson, Hashtroudi, & Lindsay, 1993). Thus, the test type does not seem be critical to whether a RES or PET pattern occurs.

A third potentially relevant difference between the paradigms—which we investigated in this study—is the format used to present the misinformation. Studies that have found RES have primarily embedded misinformation in a narrative format (e.g., Chan & LaPaglia, 2011; Chan et al., 2017; Chan et al., 2009; Chan et al., 2012; Gordon & Thomas, 2014; Thomas et al., 2010; Wilford et al., 2014). In contrast, studies showing a PET have typically embedded misinformation in cued-recall questions (e.g., Gabbert et al., 2012; Lane et al., 2001; Pansky & Tenenboim, 2011), except Huff et al. (2013) and Huff et al. (2016), who embedded misinformation via a social-contagion format (i.e., fabricated household scene recall sheets). LaPaglia and Chan (2013; see too LaPaglia & Chan, 2019) examined whether misinformation format dictates whether RES or a PET occurred in their crime-video paradigm. Importantly, misinformation presented via a written narrative produced a RES pattern, whereas misinformation presented via cued-recall questions produced a PET pattern. To explain this dissociation, LaPaglia and Chan suggested that misinformation embedded in a question encourages participants to compare their memory for the original event to what the question suggests. This comparison process may assist participants in detecting discrepant details during the misinformation phase which would subsequently aid monitoring for those details on a final test, thus yielding the PET pattern. In the narrative condition, this comparison process may be less likely given that reading the narrative does not require a particular memory response. Participants may, therefore, be more likely to report misinformation, thus yielding the RES pattern.

Building on LaPaglia and Chan (2013), our study examined whether narrative versus question misinformation format modulates whether a PET or RES occurs in the household scene paradigm. Their findings lend credence to this possibility, but there are several potentially important differences between the crime video and household scene paradigms. For example, most studies yielding RES have used video events and an initial cued-recall test (e.g., Chan et al., 2009; LaPaglia & Chan, 2013; Memon et al., 2010), whereas Huff et al.'s studies yielding a PET used static household scenes and an initial free-recall test. In addition, Huff et al.'s participants inspected fake recall sheets ostensibly completed by social others, which may have aroused suspicion, thereby facilitating source-monitoring processes during the final memory tests that advantaged the initial-test group. Manipulating misinformation formats enabled us to test whether their crossover pattern generalises to the household scene paradigm.

Following Huff et al. (2013) and Huff et al. (2016), after studying images of household scenes, half our participants completed an initial free-recall test. Following a 48-hr delay (to reduce memory for the initial event), participants were exposed to suggested scene items either in a narrative or cued-recall question format (rather than through Huff et al.'s social contagion format). Final free recall and source monitoring tests were then completed. Based on LaPaglia and Chan (2013), and the differential processing of narratives versus questions outlined above, we expected RES with the narrative format and PET with the questions format.

A secondary goal of our study was to evaluate whether initial testing might also yield a memory cost by increasing the misinformation effect for suggested details that are spontaneously reported on the initial test. Suggested items are designed to be schema-consistent, and thus may arise as false memories during an initial test. In the household-scene paradigm, suggested items are often high-expectancy items intentionally omitted from the photo (e.g., soap in a bathroom scene). Although this possibility has not been evaluated previously, participants likely recall some of these suggested items during the initial recall test. Previous research has shown that initial errors, including misinformation, often persist across subsequent tests in the absence of corrective feedback (Kang et al., 2011; Kay, 1955; Lane et al., 2001; Zhu et al., 2012). The later presentation of suggested items during the misinformation phase could reinforce those initial errors, thus increasing their likelihood of being reported on the final recall test. Indeed, this could explain why a PET pattern is sometimes absent on a final recall test, yet present on a source-monitoring test (Huff et al., 2013; Huff et al., 2016).

On this issue, Huff et al. (2013, Experiment 2) presented initial-test participants with corrective feedback after the initial test. Feedback decreased the reporting of suggested items on the final recall test, yet the PET pattern on the final recall test did not reach significance. In this study, we measured the recall rates of suggested items on the initial test, and then removed those items from the calculation of the misinformation effect on the final recall test. Recall of suggested items on the initial test should work against a PET on free recall, at least when recall instructions do not require source monitoring. Thus, we expected that the conditionalised analysis might yield a PET on the final recall test. Such a demonstration would confirm the generality of the PET pattern in the household-scene paradigm. It may also identify an important cost to initial testing, in line with the RES pattern.

To further evaluate the potency of the initial test as a source of the misinformation effect, in Experiment 2 we modified the source-monitoring test so that memory of recalling details on the initial test was included as an option. In Experiment 1 and in the source tests used in Huff et al. (2013) and Huff et al. (2016), participants did not have the option of ascribing their memory for a suggested item to their responses on the initial test. Instead, they classified the source as either the household scene photo, the misinformation phase, or neither. By adding an initial test option, we were therefore able to evaluate how often participants recollect outputting suggested items on the initial test, and the impact of that recollection when making source-monitoring decisions.

Finally, following Gordon and Thomas (2014) and Gordon et al. (2015), in Experiment 1 we measured reading times for each block of narrative or questions during the misinformation phase, to determine whether reading times were longer for blocks that contained falsely suggested items versus those that did not. In these prior studies, where narratives were used to present misinformation, reading times were longer for sentences containing suggested details relative to control sentences—particularly when participants had completed an initial test. The researchers suggested that initial testing increased the amount of processing time allocated towards suggested misleading items during the misinformation phase, resulting in increased reporting of suggested items on the final test. To gain insight into the effects of misinformation format in the household scene paradigm, Experiment 1 compared reading times for each format of blocks that suggested versus did not suggest a scene item, and we expected that reading times would be greater for blocks containing suggested items than those that did not. In Experiment 2, we equated reading times for blocks across formats to rule out reading time differences as a potential explanation for the Experiment 1 patterns.

Experiment 1

Participants

University of Calgary undergraduates participated for course credit. They were randomly assigned to the initial-test group or no-test group, and to receive misinformation either via a narrative or questions, resulting in four groups. Four participants were excluded for not complying with task instructions, and 14 participants failed to return for the second session and were excluded, leaving 148 participants for analysis (mean age = 20.70, SD = 4.40; range = 18–48). The source-test data were missing for 5 of these 148 participants.

Materials

The study materials were six digital colour images of common household scenes (*bathroom, toolbox, desk, kitchen, closet, bedroom*), taken from a normed study by Huff et al. (2016), that displayed many items (M = 23.83) commonly found in that scene. From these norms, the two most highly expected items for each scene were not included in the scene and instead served as the 12 suggested items (*soap/toothbrush, nails/screwdriver, paper/pen, knives/plates, coat/shoes, and lamp/pillow*).

A narrative and a set of cued-recall questions were constructed for each scene (see Figure 1 for an example), within which both suggested items were presented. The

Narrative	Questions
Having forgotten his cell phone[1] on the table[2], Jordan walked over to grab it.	Was the apple[14] green or red?
While he was there, he noticed that the tomorrow spot on his calendar[3] said he had an assignment due. He had completely forgotten about it.	How many CDs[5] were there?
Frustrated with himself, he sat down to complete it. He grabbed some **paper**[4] and began to think.	What colour was the case on the cellphone[1]?
After staring into space for a few minutes, he decided that some music would help his thought process. He reached for the nearest CD[5] and popped it in the laptop[6].	Were the headphones[7] coiled or tangled?
His sister yelled for him to turn his music down, so he plugged in his headphones[7] and put them in his ears.	Were the envelopes[12] front or flap side up?
He immediately felt more creative, but not for his homework. Jordan reached for the post-its[8], and started to glue them together with the glue stick[9].	Was the scotch tape[11] to the left or the right of the laptop[6]?
In an inspired moment, he linked together some paper clips[10] and taped them with scotch tape[11]. He proceeded to stick them to the rest of his creation.	Were the paperclips[10] on the post-it notes[8]?
Unsure of what he had just made, Jordan realized that he had wasted half an hour, and his assignment was no closer to completion.	Was the glue stick[9] directly on the table[2] or on the **papers**[4]?
With a sign of resignation, Jordan grabbed some envelopes[12] and began to write out an outline for his paper, using some highlighters[13] to make key points stand out.	What colour was the calendar[3]?
Finally at a point where he could concentrate, he took a bit of his apple[14], grabbed the **pen**[15], and began writing his essay.	Were the **pens**[15] with the highlighters[13] or elsewhere?

Figure 1. Sample household scene image and corresponding narrative and questions, shown in the presented order. Superscripts denote matching items across narratives and questions. Bolded items denote false items suggested to participants. Participants were not presented with superscripts or bolded items during the experiment.

narratives were vignettes in which a fictional character interacted with multiple scene items. The cued-recall questions asked participants about the proximity or features of various items. The narratives ranged from 8 to 16 sentences and were subdivided into blocks of text to equate the number of blocks of text with the number of

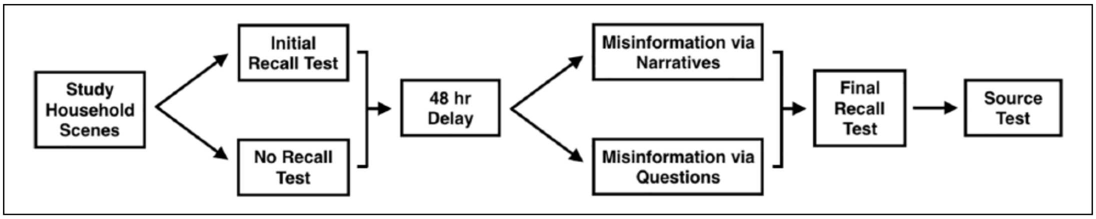

Figure 2. Study procedure for Experiments 1 and 2.

cued-recall questions for the same scene. Splitting the materials into blocks allowed us to measure the reading time for blocks that contained versus did not contain suggested items. No item was mentioned in more than one block. *Consistent items* are scene items that were also mentioned in a block (11–12 per scene). *Control items* are scene items that were not mentioned in a block (11–12 per scene). By mistake, 6 of the 144 scene items appeared only in the narrative or question block rather than in both. These items were excluded from analysis. The two suggested items for a given scene were always presented in different blocks.

The 66-item source-monitoring test included 11 items from each scene: four consistent items, four control items, two suggested items, and one related lure item that was neither in the scene nor suggested (e.g., *printer* for the desk scene). Due to the error mentioned above, two consistent items on this test were dropped from the source-test analyses. The source test consisted of four response options: "photo" if the item had been shown in the photo; "narrative" or "question" if the item had been presented in the narrative or questions, respectively; "both" if the item had been presented in both the photo and narrative/questions; or "neither."

Procedure

Figure 2 depicts the study design. Participants were tested in a computer lab in groups ranging from 10 to 30 participants. The study was conducted using Qualtrics software. Participants completed two sessions spaced 48-hr apart. In Session 1, participants were informed that they would view household scenes, and then their memory for the scenes would be tested. Consistent with Huff et al. (2013) and Huff et al. (2016), each scene was then shown for 15 s in the order listed above, preceded by the scene title. The no-test condition groups were then dismissed and instructed to return for Session 2. The initial-test groups immediately completed a 2 min free recall test for each scene in the same order. The scene title was presented, and participants typed in as many items from the scene as they could recall. The initial-test groups were then dismissed and instructed to return for Session 2.

Session 2 began with the misinformation phase. The scene title preceded each narrative or set of questions. The narrative groups read a narrative based on each household scene studied in Session 1. They read each block of text, and then clicked a "next" button to advance to the next block. The questions groups read each cued-recall question and were asked to answer it silently to themselves and then clicked a "next" button to advance to the next question. Silent responses in the questions condition were used in the questions condition to match the responses in the narrative condition. Although participants may not have answered all of the questions, their purpose was to introduce the suggested items. A maximum of 30 s was provided per block, after which the next block appeared. Reading latencies were measured from the onset of each block until the participant clicked the "next" button, or until the block advanced automatically after 30 s (in which case the reading time was recorded as 30 s). After the misinformation phase, participants completed an arithmetic filler task for 2 min.

After the filler task, participants completed the final free recall task, which was identical to the initial free recall task. They then completed the self-paced source-monitoring test for which responses were made via on-screen radio buttons.

Results

A liberal criterion was used to score free recall responses, such that misspellings and synonyms were treated as correct (as in Huff et al., 2016). For example, the word "pan" was scored as correct for the scene item "pot." Free recall and source responses for each item type were analysed using a 2 (test condition: no test vs. initial test) × 2 (misinformation format: narratives vs. questions) between-subjects analysis of variance (ANOVA). For completeness, we report the main effect of misinformation format for each ANOVA, but we do not interpret them, given we were primarily interested in whether the effect of initial testing differed for the two misinformation formats. Results were significant at the $p < .05$ level unless otherwise reported. Partial eta-squared (η_p^2) is reported as a measure of effect size for ANOVAs and Cohen's d for all t-tests. A sensitivity analysis using G*POWER (Erdfelder, Faul, & Buchner, 1996), indicated that our sample size had sufficient power (.80) to detect medium-sized effects (e.g., Cohen's $d = .50$ or larger) for main effects and interactions.

Table 1. Mean (*SD*) proportion of items recalled on the final recall test in Experiments 1 and 2 as a function of item type, misinformation format (narratives vs. questions), and initial test condition (no test vs. initial test).

Experiment/item type	Narratives		Questions	
	No test	Initial test	No test	Initial test
Experiment 1				
N	35	38	35	40
Consistent	.31 (.12)	.44 (.14)	.36 (.09)	.46 (.11)
Control	.08 (.05)	.11 (.07)	.09 (.05)	.18 (.08)
Suggested/conditional	.39 (.20)	.45/.36 (.18/.16)	.48 (.19)	.48/.37 (.19/.20)
Experiment 2				
N	35	33	36	37
Consistent	.35 (.12)	.43 (.13)	.40 (.14)	.41 (.12)
Control	.12 (.06)	.12 (.09)	.11 (.06)	.15 (.06)
Suggested/conditional	.46 (.19)	.46/.34 (.17/.17)	.52 (.22)	.47/.36 (.22/.20)

SD: standard deviation.

Non-significant effects were further tested using a Bayesian estimate of the strength of evidence supporting the null hypothesis (Masson, 2011; Wagenmakers, 2007). This analysis compares two models in which one assumes an effect and the other assumes a null effect. The Bayesian analysis yields a probability estimate that the null effect is retained: A *p*-value termed p_{BIC} (Bayesian Information Criterion). This analysis is sensitive to sample size and therefore is more informative regarding null effects than null-hypothesis significance testing.

Initial free recall test

The mean proportion of scene items correctly recalled on the initial test was 0.27 per scene ($SD=.08$), and the mean proportion of suggested items falsely recalled was 0.23 per scene ($SD=.46$). There were no differences in the rates of correct or suggested item recall for participants who later received narratives versus questions, $t\text{s} < 1.04$, $p_{BIC}\text{s} > .87$.

Final free recall test

Table 1 provides the mean proportion of items recalled on the final recall test as a function of item type, misinformation format, and initial test condition. We first report analyses for consistent and control items. We then focus on our primary interests, namely, the influence of initial testing on memory for suggested items, and whether the direction of this influence depended on misinformation format.

Consistent items. Initial testing enhanced free recall of consistent items (i.e., scene items later mentioned in the misinformation phase) relative to the no-test condition (.45 vs. .34), $F(1, 144)=35.10$, mean square error (MSE)=.01, $\eta_p^2=.20$, a retrieval-practice effect (Roediger & Karpicke, 2006). Marginally fewer consistent items were recalled in the narrative versus questions condition (.38 vs. .41), $F(1, 144)=2.96$, MSE=.01, $p=.09$, $\eta_p^2=.02$, $p_{BIC}=.78$. The interaction was not significant, $F<1$, $p_{BIC}=.91$.

Control items. Initial testing also increased correct recall of control items (i.e., scene items not mentioned during the misinformation phase) relative to the no-test condition (.15 vs. .09), $F(1, 144)=25.55$, MSE=.01, $\eta_p^2=.15$. Control items were less likely to be recalled in the narrative versus questions format condition (.10 vs. .14), $F(1, 144)=13.12$, MSE=.01, $\eta_p^2=.08$. These effects were qualified by a significant interaction, $F(1, 144)=8.18$, MSE=.01, $\eta_p^2=.05$. Follow-up *t*-tests revealed that initial testing significantly increased recall of control items in the questions condition (.18 vs. .09), $t(73)=3.98$, standard error of the mean (SEM)=.01, $d=0.91$, but not in the narratives condition (.09 vs. .08), $t<1$, $p_{BIC}=.88$.

Suggested items. Turning to the misinformation effect, recall of suggested items was equivalent in the test and no-test conditions (.46 vs. .44)—a null effect of test condition, $F<1$, $p_{BIC}=.89$. Marginally fewer suggested items were recalled when misinformation was embedded in a narrative rather than in questions (.42 vs. .48), $F(1, 144)=3.40$, MSE=.04, $p=.07$, $p_{BIC}=.70$, $\eta_p^2=.02$. The interaction was not reliable, $F(1, 144)=1.08$, MSE=.04, $p=.30$, $p_{BIC}=.88$.

As described above, 23% of suggested items were spontaneously recalled on the initial test, which was not surprising given these were high-expectancy items. On the final recall test, participants may have recollected having output these items on the initial recall test and used this recollection as the basis for outputting them again. Taking an initial test may encourage participants to generate their own misinformation, at least for plausible, schema-consistent details (e.g., Brewer & Treyens, 1981). Therefore, we next isolated the effect of initial testing on recall of suggested items that were not reported on the initial test. To this end, we conducted a conditionalised analysis in which suggested items reported on the initial test were

Table 2. Experiment 1: mean (*SD*) proportion of items attributed to source-monitoring test options by item type, misinformation format, and initial test condition.

Item type	Attribution	Narratives		Questions	
		No test	Initial test	No test	Initial test
Consistent	Photo	.24 (.13)	.14 (.10)	.28 (.11)	.18 (.12)
	Both	.41 (.27)	.52 (.18)	.41 (.19)	.43 (.17)
	Narrative/questions	.18 (.12)	.23 (.14)	.17 (.11)	.29 (.17)
	Neither	.18 (.15)	.11 (.11)	.14 (.14)	.10 (.10)
Control	Photo	.42 (.20)	.39 (.19)	.50 (.19)	.43 (.16)
	Both	.11 (.10)	.10 (.09)	.03 (.04)	.04 (.05)
	Narrative/questions	.05 (.07)	.06 (.05)	.02 (.03)	.02 (.03)
	Neither	.40 (.19)	.45 (.18)	.44 (.18)	.51 (.15)
Related Lures	Photo	.35 (.23)	.20 (.20)	.33 (.21)	.20 (.22)
	Both	.10 (.13)	.06 (.09)	.02 (.05)	.02 (.05)
	Narrative/Questions	.05 (.09)	.06 (.09)	.02 (.05)	.02 (.05)
	Neither	.49 (.23)	.67 (.22)	.63 (.21)	.74 (.22)
Suggested	Photo	.33 (.17)	.18 (.14)	.40 (.14)	.25 (.19)
	Both	.45 (.18)	.51 (.22)	.38 (.19)	.36 (.20)
	Misinformation effect	**.77 (.23)**	**.69 (.20)**	**.78 (.20)**	**.61 (.27)**
	Narrative/questions	.13 (.14)	.19 (.15)	.09 (.10)	.26 (.23)
	Neither	.09 (.13)	.13 (.14)	.13 (.13)	.13 (.14)

SD: standard deviation.
The "both" source option reflects source retrievals from both the photos and the narrative/questions. For suggested items, the misinformation effect row (bold) is the sum of the photo and both rows of misattributions. For suggested items, the correct response was the "narrative/questions" option. For consistent items, the correct response was the "both" option. For control items, the correct response was the "photo" option. For related lures, the correct response was the "neither" option.

removed from analysis of the final recall test (i.e., by subtracting them from *both* the numerator and the denominator). The ANOVA on conditionalised suggested item recall (see Table 1) now revealed a PET: Initial testing reduced false recall of suggested items that were not recalled on the initial test, relative to the no-test condition (.37 vs. .44), $F(1, 144)=5.40$, MSE=.04, $\eta_p^2=.04$. Neither the main effect of misinformation format nor the interaction reached significance, $F(1, 144)=2.81$, MSE=.04, $p=.10$, $p_{BIC}=.74$, and $F(1, 144)=1.52$, MSE=.04, $p=.22$, $p_{BIC}=.85$.

Source monitoring test

Table 2 provides the mean proportion of source attributions as a function of item type, misinformation format, and initial test condition.

Consistent items. For consistent items, the correct response on the source test was selection of the "both" option, indicating items were found in both the photos and the narrative/questions sources (Huff et al., 2013; Huff et al., 2016). The ANOVA revealed a main effect of test condition, such that initial testing increased the rate of correct "both" attributions relative to the no-test condition (.47 vs. .41), $F(1, 139)=4.31$, MSE=.03, $\eta_p^2=.03$. The effect of misinformation format was not significant, $F(1, 139)=2.23$, MSE=.03, $p=.14$, $p_{BIC}=.76$, nor was the interaction, $F(1, 139)=2.60$, MSE=.03, $p=.11$, $p_{BIC}=.76$.

Control items. For control items, "photo" was the only correct source option. Here, the ANOVA revealed a reversed testing effect, such that correct source attributions for control items were *less likely* after an initial test than in the no-test condition (.41 vs. .47), $F(1, 139)=5.13$, MSE=.04, $\eta_p^2=.04$. Interestingly, previous studies showing a PET pattern also reported this reversal (Huff et al., 2013; Huff et al., 2016); we consider this negative consequence of initial testing in section "General Discussion." The main effect of misinformation format failed to reach significance, $F(1, 139)=2.20$, MSE=.04, $p=.14$, $p_{BIC}=.80$, as did the interaction, $F<1$, $p_{BIC}=.92$.

Related lures. For related lure items, "neither" was the correct source attribution. The ANOVA showed a main effect of test condition; correct attributions were more likely in the initial-test versus no-test condition (.71 vs. .56), $F(1, 139)=14.70$, MSE=.05, $\eta_p^2=.10$. There was also a main effect of misinformation format, which indicated that "neither" attributions were less likely in the narrative than questions condition (.58 vs. .69), $F(1, 139)=7.48$, MSE=.05, $\eta_p^2=.05$. The interaction was not significant, $F(1, 139)=1.05$, MSE=.05, $p=.31$, $p_{BIC}=.87$.

Table 3. Experiment 1: mean (*SD*) reading times (in s) during the misinformation phase as a function of block type (no suggested item vs. suggested item), misinformation format (narratives vs. questions), and initial test condition (no test vs. initial test).

Block type	Narratives		Questions	
	No test	Initial test	No test	Initial test
No suggested item block	5.14 (1.59)	6.14 (2.46)	4.77 (1.47)	5.20 (3.09)
Suggested item block	4.54 (1.37)	5.64 (2.51)	4.67 (1.40)	5.97 (3.79)

SD: standard deviation.

Suggested items. The misinformation effect was computed as the proportion of suggested items that were incorrectly attributed to the scenes either via a "photo" or "both" response options. A main effect of test condition was found: initial testing (vs. no testing) reduced the proportion of suggested items that were incorrectly attributed to the scenes (.65 vs. .78), $F(1, 139)=11.67$, $MSE=.05$, $\eta_p^2=.08$. Thus, replicating Huff et al. (2013) and Huff et al. (2016), source judgements showed a PET pattern. The effect of misinformation format and the interaction were not significant, $F<1$, $p_{BIC}=.88$, and $F(1, 139)=1.32$, $MSE=.05$, $p=.25$, $p_{BIC}=.86$.

Correct attributions for suggested items (i.e., to the narratives/questions source) were also analysed. Correct source attributions were more likely in the initial-test (vs. no-test) condition (.22 vs. .11), $F(1, 139)=16.47$, $MSE=.03$, $\eta_p^2=.11$. The effect of misinformation format was not significant, $F<1$, $p_{BIC}=.91$. However, the interaction was significant, $F(1, 139)=4.39$, $MSE=.03$, $\eta_p^2=.03$. Follow-up tests revealed that the increase in correct source attributions for suggested scene items in the initial test (vs. no test) condition was significant in the questions condition (.26 vs. .09), $t(71)=3.99$, $SEM=.04$, $d=0.95$, but not in the narratives condition (.19 vs. .13), $t(68)=1.56$, $SEM=.05$, $p=.12$, $p_{BIC}=.69$.

Reading times during misinformation phase

Table 3 reports the mean reading times for the narrative and question blocks during the misinformation phase as a function of whether the block contained a suggested detail. A 2 (test condition: no test vs. initial test) × 2(misinformation format: narratives vs. questions) × 2 (block type: no suggested item vs. suggested item) mixed ANOVA was used to analyse average reading times. The main effects of misinformation format and block type were not significant, $F<1$, $p_{BIC}=.91$, $F(1, 139)=2.43$, $MSE=.35$, $p=.12$, $p_{BIC}=.78$. A significant main effect of test condition was found: Reading times were longer following an initial test than no initial test (5.74 s vs. 4.78 s), $F(1, 139)=5.78$, $MSE=11.32$, $\eta_p^2=.04$. This effect was qualified by a significant three-way interaction, $F(1, 139)=7.38$, $MSE=.35$, $\eta_p^2=.05$. Follow-up tests revealed that this interaction was due to initial testing resulting in longer reading times in all conditions except for no suggested item blocks in the question format condition, in which reading times following initial testing were nonsignificantly slower (5.20 s vs. 4.77 s), $t<1$, $p_{BIC}=.87$.

Discussion

Using a crime-video paradigm, LaPaglia and Chan (2013) found that initial testing yielded a RES pattern when misinformation was introduced via a narrative but a PET pattern when misinformation was introduced via questions. In Experiment 1, we similarly predicted this pattern using the household-scene paradigm (Huff et al., 2013; Huff et al., 2016), but instead obtained a PET pattern of similar magnitude for both narrative and question formats. The PET pattern on the source-monitoring test replicated Huff et al.'s earlier studies where misinformation was introduced through a social format (fake recall sheets from "other participants"). Further replicating those studies, initial testing resulted in fewer attributions of suggested items to the scenes, and it also increased correct attributions of these items to the narrative or question source. The latter effect was significant only for the questions format, consistent with the possibility that reading a narrative may be more likely to encourage participants to attempt to distinguish the source of their memory for each detail.

Although a PET pattern was absent in our overall recall analysis, it was significant in a conditional analysis which omitted suggested items recalled on the initial test. This PET was also not significantly influenced by misinformation format. Huff et al. (2016) obtained a significant PET on overall free recall, but only when a 48-hr delay preceded the misinformation and final test phases (as was the case in Experiment 1). Moreover, some of Huff et al.'s participants took two initial tests, and some of the suggested items were presented four times among the fake recall sheets. In Experiment 1, there was only one initial test and each suggested item was presented only once. These differences likely contributed to the PET being absent in our overall analysis.

Huff et al. (2013) suggested that the absence of a PET on free recall was due to this task not requiring retrieval of source details. However, our conditionalised analyses suggest a PET in recall might be masked by the frequent output of high expectancy suggested items on the initial test. Confirming this possibility, ANOVAs performed on the

Table 4. Mean (*SD*) proportion of conditionalised suggested items recalled on the final recall test by prior experiment, misinformation exposure, and initial test condition.

Experiment/misinformation exposures	No test	Initial test
Huff, Davis, and Meade (2013, Experiment 1)		
One Exposure	.17 (.21)	.08 (.12)
Four Exposures	.38 (.32)	.31 (.28)
Huff, Weinsheimer, and Bodner (2016, one initial test, no delay)		
One Exposure	.42 (.31)	.17 (.22)
Four Exposures	.54 (.26)	.29 (.26)
Huff et al. (2016, one initial test, 48-hr delay)		
One Exposure	.40 (.25)	.11 (.14)
Four Exposures	.65 (.27)	.30 (.29)

SD: standard deviation.
The Huff et al. (2013) means were based on recall of both high and low expectancy suggested items, whereas Huff et al. (2016) only used high expectancy suggested items.

conditionalised means (see Table 4, averaged across one and four misinformation exposure conditions) of Huff et al. (2013, Experiment 1), Huff et al. (2016, one initial test, no delay), and Huff et al. (2016, one initial test, 48-hr delay), each revealed a PET on recall: $F(1, 70)=5.87$, $MSE=.04$, $\eta_p^2=.08$, $F(1, 70)=26.75$, $MSE=.08$, $\eta_p^2=.27$, and $F(1, 70)=55.86$, $MSE=.07$, $\eta_p^2=.44$, respectively.

These conditionalised analyses establish that in addition to yielding memory benefits, initial testing can also yield memory costs. Specifically, information falsely recalled on an initial test is more likely to be reported later. Experiment 2 takes a closer look at this issue using a modified source test. On the benefits side, the reduction in the misinformation effect after an initial test was complimented by a retrieval-practice effect on correct recall of scene items (Roediger & Karpicke, 2006). Thus, initial free-recall testing can enhance overall memory accuracy, as reported by Huff et al. (2016). This testing benefit was absent on cued-recall accuracy in LaPaglia and Chan's (2013) crime video paradigm, even when a PET was obtained following misleading questions.

Finally, we also analysed reading times for blocks of narratives and questions as a function of initial test condition and misinformation format. Consistent with Gordon and Thomas (2014), reading times were generally longer after participants completed an initial test. Gordon and Thomas argued that longer reading latencies reflect increased attentional processing of suggested information. We similarly found longer reading times in the initial test condition, but contrary to Gordon and Thomas' findings, these longer reading times were associated with a PET rather than RES pattern. In the household-scene paradigm, longer reading times may be reflective of a discrepancy detection process in which participants notice the presence of a non-studied item. Consistent with this interpretation, Tousignant, Hall, and Loftus (1986) found that the detection of misinformation contradictions was associated with longer reading times during misinformation exposure. Assuming that participants are more successful at detecting misinformation in the narratives and questions following initial testing, monitoring for these items may increase on the final test, thereby reducing the misinformation effect.

Experiment 2

Experiment 1 confirmed that initial testing can have both positive and negative consequences on memory. A PET effect emerged in free recall when suggested items spontaneously recalled on the initial test were removed. Our suggested items were all high-expectancy items for their scenes, akin to critical lures in the Deese–Roediger–McDermott (DRM) paradigm (Roediger & McDermott, 1995), and indeed were often recalled on the initial test. An important possibility left unaddressed by Experiment 1 is whether participants who took an initial test sometimes reported suggested items on the final test because they recollected having reported them on the initial test (rather than because they recollected having seen them in the photos). The overall analysis of recall may have failed to show a PET because participants did not monitor for whether they recollected seeing the suggested items in the photos versus having reported them on the initial recall test. In contrast, the source test required participants to consider the source of their memory for each item, and in doing so, may have helped them avoid attributing suggested items to the photos. The primary goal of Experiment 2 was to evaluate this possibility. To this end, participants in the initial-test condition were offered an additional "initial test" source option on the source-monitoring test. This option provided us with a measure of how often suggested items were attributed to the initial-test source.

The secondary goal of Experiment 2 was to provide a second opportunity to evaluate the effects of misinformation format on memory for suggested items. Moreover, this time, rather than allowing self-paced reading, each block of text was presented for the same duration to control the amount of time spent processing blocks containing versus not containing suggested items for both formats. Although reading times for blocks of text did not differ significantly across the narrative versus question formats in Experiment 1, participants may have traded off speed versus effort in processing more for one format than another. Preventing participants from spending longer processing blocks containing suggested details will also help establish the generality of the effects of misinformation format.

Method

Participants

Additional University of Calgary undergraduates not tested in Experiment 1 were randomly assigned to the same four groups used in Experiment 1. Four participants were excluded for failure to follow instructions, and 25 failed to return for the second session, leaving 141 participants for analysis (mean age = 20.40, *SD* = 3.73; range = 18–41).

Materials

The materials were the same as Experiment 1 except the items from Experiment 1 that were mistakenly excluded from one or the other format were properly included.

Procedure

With two exceptions, the procedure was the same as Experiment 1. First, each narrative or question block during the misinformation phase was presented for 8 s, reflecting the mean reading time per block plus one *SD*, based on the narrative condition means in Experiment 1. Second, the source-monitoring test was modified such that an "initial test" option was added to enable participants to indicate that they remembered reporting the item on the initial test. To accommodate this addition, participants individually selected all of the sources that they recollected for a given item (i.e., photo and/or initial test and/or narrative/questions).

Results

Analysis followed Experiment 1, except as noted. A sensitivity analysis similarly indicated that our sample size had sufficient statistical power (.80) to detect medium-sized effects (Cohen's $d = 0.52$ or larger) for all main effects and interactions.

Initial free recall test

On average, participants who received the initial test correctly recalled 0.25 (*SD* = 0.06) of the scene items and falsely recalled 0.25 (*SD* = 0.18) of the suggested items. Initial test recall was similar whether participants later received narratives or questions, $t\text{s} < 1.51$, $p_{\text{BIC}}\text{s} > .78$.

Final free recall test

Table 1 provides the mean proportion of items recalled on the final recall test as a function of item type, misinformation format, and initial test condition.

Consistent items. Initial testing enhanced final recall of consistent items relative to the no-test condition (.42 vs. .37), $F(1, 137) = 5.51$, MSE = .02, $\eta_p^2 = .04$. Recall of consistent items did not differ between the questions and narratives conditions (.40 vs. .39), $F < 1$, $p_{\text{BIC}} = .91$, and the interaction did not attain significance, $F(1, 137) = 2.72$, MSE = .02, $p = .10$, $p_{\text{BIC}} = .75$.

Control items. Initial testing marginally enhanced final recall of control items over the no-test condition (.14 vs. .11), $F(1, 137) = 3.57$, MSE = .01, $p = .06$, $\eta_p^2 = .03$, $p_{\text{BIC}} = .66$. The effect of misinformation format was not significant, $F < 1$, $p_{\text{BIC}} = .85$. Unlike in Experiment 1, where recall was enhanced only in the questions condition, the interaction was not significant, $F < 1$, $p_{\text{BIC}} = .77$.

Suggested items. As in Experiment 1, overall recall of falsely suggested items was equivalent between the initial-test and no-test conditions (.46 vs. .49), $F < 1$, $p_{\text{BIC}} = .89$, demonstrating neither a PET nor RES pattern. Recall of suggested items did not differ as a function of misinformation format, $F(1, 137) = 1.22$, MSE = .04, $p = .27$, $p_{\text{BIC}} = .86$, and the interaction was also not significant, $F < 1$, $p_{\text{BIC}} = .90$. However, as in Experiment 1, the conditionalised analysis revealed a robust PET: initial testing reduced recall of suggested items relative to the no-test condition (.35 vs. .49), $F(1, 137) = 19.17$, MSE = .04, $\eta_p^2 = .12$; here the main effect of misinformation format and the interaction remained nonsignificant, $F(1, 137) = 1.56$, MSE = .04, $p = .21$, $p_{\text{BIC}} = .84$, and $F < 1$, $p_{\text{BIC}} = .91$.

Source monitoring test

Table 5 provides the mean proportion of source attributions as a function of item type, misinformation format, and initial test condition.

Consistent items. For consistent items, we scored a response as correct if the participant selected both the "photo" and "narratives/questions" options. In the initial-test condition, we scored a response as correct regardless of whether they also selected the "initial test" source option (i.e., we summed rows 2 and 4 in Table 5). Unlike Experiment 1, correct attributions for consistent items were *lower* in the initial-test versus no-test condition (.29 vs. .40), $F(1, 137) = 10.81$, MSE = .04, $\eta_p^2 = .07$. The inclusion of the

Table 5. Experiment 2: mean (*SD*) proportion of items attributed to source-monitoring test options by item type, misinformation format, and initial test condition.

Item type	Attribution	Narratives		Questions	
		No test	Initial test	No test	Initial test
Consistent	Photo	.27 (.16)	.13 (.12)	.29 (.17)	.15 (.12)
	Photo + narrative/questions	.38 (.19)	.15 (.15)	.42 (.21)	.13 (.14)
	Photo + initial test	–	.08 (.10)	–	.07 (.08)
	Photo + narrative/questions + initial test	–	.14 (.15)	–	.16 (.14)
	Narrative/questions	.24 (.13)	.27 (.12)	.20 (.14)	.30 (.11)
	Initial test	–	.04 (.10)	–	.03 (.09)
	Narrative/questions + initial test	–	.02 (.05)	–	.01 (.03)
	Neither	.11 (.12)	.18 (.13)	.09 (.09)	.14 (.10)
Control	Photo	.46 (.14)	.21 (.17)	.41 (.19)	.22 (.16)
	Photo + narrative/questions	.08 (.06)	.04 (.05)	.02 (.03)	.01 (.03)
	Photo + initial test	–	.14 (.15)	–	.15 (.13)
	Photo + narrative/questions + initial test	–	.03 (.04)	–	.02 (.03)
	Narrative/questions	.03 (.04)	.05 (.09)	.04 (.07)	.05 (.07)
	Initial test	–	.04 (.08)	–	.03 (.06)
	Narrative/questions + initial test	–	.00 (.01)	–	.00 (.01)
	Neither	.44 (.17)	.50 (.19)	.53 (.18)	.51 (.15)
Related lure	Photo	.30 (.20)	.11 (.15)	.28 (.24)	.16 (.19)
	Photo + narrative/questions	.06 (.09)	.02 (.06)	.02 (.06)	.00 (.03)
	Photo + initial test	–	.02 (.06)	–	.02 (.06)
	Photo + narrative/questions + initial test	–	.02 (.06)	–	.00 (.03)
	Narrative/questions	.05 (.08)	.08 (.09)	.02 (.06)	.03 (.08)
	Initial test	–	.03 (.07)	–	.01 (.06)
	Narrative/questions + initial test	–	.01 (.24)	–	.00 (.00)
	Neither	.58 (.19)	.73 (.24)	.68 (.23)	.76 (.21)
Suggested	Photo	.34 (.18)	.18 (.20)	.36 (.18)	.16 (.16)
	Photo + narrative/questions	.39 (.21)	.12 (.16)	.38 (.18)	.14 (.15)
	Photo + initial test	–	.08 (.17)	–	.10 (.14)
	Photo + narrative/questions + initial test	–	.16 (.18)	–	.14 (.14)
	Misinformation effect	**.73 (.20)**	**.54 (.16)**	**.75 (.20)**	**.55 (.20)**
	Narrative/questions	.15 (.13)	.24 (.19)	.16 (.15)	.25 (.16)
	Narrative/questions + initial test	–	.02 (.05)	–	.01 (.03)
	Initial test	–	.05 (.09)	–	.02 (.15)
	Neither	.12 (.14)	.15 (.15)	.10 (.10)	.17 (.10)

SD: standard deviation.
The "both" source option reflects source retrievals from both the photos and the narrative/questions. The "initial test" source option reflects retrievals from items recalled in initial test. For suggested items, the misinformation effect row (bold) is the sum of the photo and both rows of misattributions. For suggested items, the correct response was the "narrative/questions" option. For consistent items, the correct response was the "both" option. For control items, the correct response was the "photo" option. For related lures, the correct response was the "neither" option.

additional initial test source option undid the benefits of initial testing here. We speculate on why this occurred in our section "General Discussion." The effect of misinformation format and the interaction were not significant, $F\text{s} < 1$, $p_{\text{BIC}}\text{s} > .90$.

Control items. For control items, "photo" was the correct response in the no-test condition, whereas the sum of "photo" and "photo + initial test" was computed as correct in the initial-test condition. Consistent with Experiment 1, a reversed testing effect was found on controls in which correct source attributions were lower in the initial-test condition than the no-test condition (.36 vs. .44), $F(1, 137) = 7.82$, MSE = .03, $\eta_p^2 = .05$. The effect of misinformation format was not significant, $F < 1$, $p_{\text{BIC}} = .92$, nor was the interaction, $F(1, 137) = 2.46$, MSE = .03, $p = .12$, $p_{\text{BIC}} = .77$.

Related lures. The correct source attribution for related lures was "neither." Initial testing increased correct source attributions for related lures relative to the no-test condition (.75 vs. .63), $F(1, 137) = 9.98$, MSE = .05, $\eta_p^2 = .07$, as in Experiment 1. Correct attributions were marginally more common in the questions versus narratives condition

(.72 vs. .65), $F(1, 137) = 3.09$, MSE = .05, $p = .08$, $\eta_p^2 = .02$, $p_{BIC} = .71$. The interaction was not significant, $F < 1$, $p_{BIC} = .89$.

Suggested items. The misinformation effect was computed as the proportion of suggested items that were incorrectly attributed to the "photo" source (see Table 5). For the no initial-test group, this effect was the sum of "photos" and "photos + narratives/questions" rows in Table 5. For the initial-test group this was the sum of all the rows of Table 5 that included the "photos" option. As can be seen in the "initial test" row of Table 5, suggested items were seldom attributed solely to this source, although this source was often endorsed along with other source test options. Thus, participants often recalled outputting suggested items on the initial test. As in Experiment 1, initial testing reduced the misinformation effect relative to the no-test condition (.54 vs. .74), $F(1, 137) = 28.23$, MSE = .05, $\eta_p^2 = .17$. The effect of misinformation format was not significant. The interaction was also not significant, Fs < 1, p_{BIC}s $> .91$, providing further evidence that the format of misinformation delivery did not alter the presence of a PET pattern.

Finally, replicating Experiment 1, correct attributions for suggested items (i.e., to the narratives/questions source) were more likely in the initial-test (vs. no-test) condition (.25 vs. .15), $F(1, 137) = 12.59$, MSE = .03, $\eta_p^2 = .08$. Neither the effect of misinformation format nor the interaction were significant, Fs < 1, p_{BIC}s $> .92$.

Discussion

Replicating Experiment 1, completing an initial test reduced conditionalised false recall and source misattributions relative to the no-test condition. This PET pattern was again similar whether misinformation was embedded in narratives or in misleading questions (cf. LaPaglia & Chan, 2013). Inclusion of the "initial test" option on the source-monitoring test provided new insight into participants' source knowledge for suggested items. Specifically, participants who took an initial test often later indicated having recalled suggested items on that test. Thus, the conditionalised analyses we report appropriately gauge the misinformation effect because they exclude suggested items reported on the initial test. The high rate of final recall of suggested items recalled on the initial test masked a PET effect on overall recall, highlighting the importance of taking participants' initial test responses into consideration when evaluating the effect of initial tests on the accuracy of eyewitness memory.

General discussion

Using a crime-event paradigm, LaPaglia and Chan (2013) found that asking participants to take an initial memory test prior to misinformation exposure via narratives produced a RES pattern (an increase in the misinformation effect), whereas taking a test prior to misinformation exposure via misleading cued-recall questions produced a PET pattern (a decrease in the misinformation effect). In two experiments, we examined whether a similar dissociation occurs in Huff et al.'s (2013) and Huff et al.'s (2016) household-scene paradigm, which to date has consistently yielded a PET effect. Both our experiments showed a PET pattern that was similar in size whether the misinformation format was presented as narratives or as questions. As discussed below, the effects of initial testing in this paradigm have not been shown to be critically influenced by the means through which misinformation is presented. Identifying other potential predictors of whether RES or a PET occurs is thus a key area for future research—one that will inform guidelines for the appropriate use of initial testing with eyewitnesses.

A PET pattern on conditionalised free recall

A PET effect was obtained on free recall in both of our experiments when suggested items recalled on the initial test were excluded. These conditionalised analyses also revealed a PET on recall in all three previous data sets in which a PET on recall was otherwise absent (see Table 4). Unsurprisingly, participants who recall some suggested items on an initial test show a tendency to report them again on a final test, and this tendency works against a PET. Consistent with this explanation, in Experiment 2 we provided participants with an initial test option on the source test. This option was endorsed for 31% of suggested items in the narrative condition and for 27% of suggested items in the questions condition. Thus, on the final recall test, participants frequently reported items that they later attributed, at least in part, to a memory of reporting those items on the initial recall test.

A potential concern with our conditionalised analyses is that only the initial test group's final recall can be conditionalised on initial test responses—even though the no initial test group would have falsely encoded high-expectancy items as often. We acknowledge this limitation, but note that removing suggested items reported on the initial test from the initial test group does not de facto create a PET pattern. Suggested items recalled on the initial test were excluded from both the numerator and the denominator when computing the proportion of recalled suggested items. Potentially, then, the misinformation effect in the conditionalised analysis could have increased, decreased, or stayed the same as in the overall recall analysis and thus could have yielded either a PET, RES, or a null difference relative to the no initial test condition.

The conditionalised analyses reveal an important cost of initial testing, in that errors reported initially are likely to be recalled on a subsequent test. In most misinformation paradigms, suggested items are carefully selected to be schematically consistent with the study events, and thus generate high intrusion rates (e.g., Meade & Roediger, 2002; Roediger, Meade, & Bergman, 2001). In the present experiments, 23%–25% of to-be-suggested items were reported on the initial recall test. These high rates were enough to mask a PET pattern in our overall recall analyses. Thus, a lesson from our study is that initial testing of a witness's memory for an event can backfire, increasing a witness's susceptibility to suggestions about event details that did not occur but which given their plausibility were reported on the initial test. Stated differently, the protective benefits of initial testing may be reduced for suggested items that are plausible for a given eyewitness event.

Influences of an initial test on source monitoring

When participants were required to specify the source of their retrievals (i.e., on our source test), we consistently found a PET pattern in which initial testing decreased attributions of suggested items to the household scenes. Importantly, this pattern was found without conditionalising the source data based on initial recall of suggested items. On the contrary, the effects of initial testing on source judgements for consistent items (i.e., those presented in both the photo and narratives/questions) were inconsistent across experiments. Initial testing increased correct source attributions for consistent items in Experiment 1, but decreased correct source attributions in Experiment 2 (i.e., a reversed testing effect). The reversed testing effect coincided with the addition of an initial test response option on the source test. Speculatively, participants in Experiment 2 may have been more likely to attribute consistent items to the initial test than to the photos, perhaps due to their temporal contiguity and confusability in Session 1 (see Lindsay, 1990). Correct source attributions for control items (those presented only in the photo) were also lower following initial testing (as in Huff et al., 2013; Huff et al., 2016). The low rates of correct recall on the initial test may have lead the initial-test group to adopt a more conservative response criterion on the source test than the no-test group, reducing their willingness to claim they saw the control items in the photo. Consistent with this suggestion, initial testing led to higher rates of "neither" responses to related lure items (see Tables 2 and 5; see also Huff et al., 2016). This intriguing possibility awaits testing.

What determines the RES versus PET pattern?

An important unresolved issue is why initial testing yields a PET in the household scene paradigm but RES in the crime-video paradigm. A first salient difference between the paradigms is the type of event. For example, LaPaglia and Chan (2013) used an episode of *Flashpoint* as their event, which provided a cohesive narrative at study, whereas we employed a set of photos of household scenes that did not include any narrative content. Perhaps, then, initial testing for events with a narrative structure increases susceptibility to the misinformation effect, whereas initial testing for events without a narrative structure decreases it. In contrast, however, Pansky and Tenenboim (2011) obtained a PET pattern after participants studied a set of pictures accompanied by a narrative. Thus, a PET pattern can occur whether or not the initial event includes a cohesive narrative structure.

A second potential key to whether RES or a PET occurs, suggested by Chan et al. (2017), is whether the *misinformation* (rather than the *event*) is presented as a cohesive narrative. LaPaglia (2013; Experiment 3A) manipulated whether participants were presented with a coherent or disjointed narrative. The RES occurred in the standard narrative condition but was absent when the narrative was disjointed (though a PET was not obtained). Participants who read a disjointed narrative were also more likely to claim that they compared the narrative to the original event during the misinformation phase, which would work to improve source memory accuracy.

Extending this work, LaPaglia and Chan (2019) compared the effects of initial testing for narratives (Experiment 1) and questions (Experiment 2) that either did or did not reinstate contextual information. For both formats, RES occurred when cohesive contextual information from the initial event was reinstated during the misinformation phase, but not when few contextual details were presented (again, a PET was not obtained). For questions, the same pattern occurred except here the PET approached significance in the without-context condition. Thus, Lapaglia and Chan concluded that contextual support rather than narrative versus questions format dictates whether RES occurs. However, contextual support did not dictate whether a PET pattern was found across the two formats.

The narratives we developed for our household-scene paradigm may have been harder to follow (i.e., more disjointed) than a typical narrative, given the fictional characters had to interact with multiple items for each scene. Nonetheless, there was certainly more narrative structure in our narrative condition than in our questions condition, yet the magnitude of the PET pattern was consistent between formats (cf. LaPaglia & Chan, 2019). A remaining possibility is that the occurrence of RES may depend on having an overlapping narrative structure between the event and the misinformation phase, as was the case in LaPaglia and Chan (2019). In our narrative condition, there was still a disconnect between our event (no narrative structure) and our misinformation phase (narrative structure). This could be explored in future research, for example, by adding auditory narratives during the presentation of the household scenes (see Pansky & Tenenboim, 2011) to ensure a narrative context at the time of study.

A third potential factor dictating whether a RES or PET occurs is the type of suggested details. In the household-scene paradigm, the suggested items were always additive (i.e., suggested items that supplemented the items in the scenes), whereas suggested details used in most RES studies (e.g., Chan et al., 2009; LaPaglia & Chan, 2013; Thomas et al., 2010) contradicted the original event (e.g., a hypodermic syringe vs. a chloroform rag). However, Gordon et al. (2015) found RES for both additive and contradictory misinformation introduced via narratives, and in fact, the RES effect was stronger for additive items. However, the additive and contradictory misleading details were different items. Therefore, the reported pattern could reflect item effects rather than differences between additive and contradictory misinformation.

Recently, Huff and Umanath (2018) compared additive and contradictory misinformation types while counterbalancing item differences across misinformation types. Although they did not examine initial testing effects, misinformation rates were nearly twice as large for additive than contradictory misinformation. The authors suggested that contradictory information was more likely to be detected during the misinformation phase and subsequently rejected at test. Indeed, instructing participants to report detected misinformation during the misleading question phase reduced contradictory misinformation, but had no influence on additive misinformation, relative a control group who were not instructed to report detected misinformation.

Finally, a fourth potential influence may be the amount of time participants spend processing misinformation based on whether they took an initial test. As reviewed above, Gordon and Thomas (2014) and Gordon et al. (2015) showed that initial testing (vs. no testing) increased reading times when misinformation was embedded in either questions or narratives, relative to items that did not contain misinformation. Initial testing was deemed to have increased attention to the misleading items, thus enhancing their encoding. In Experiment 1, we similarly found that initial testing led to an increase in reading times during the misinformation phase (see also Tousignant et al., 1986). Yet, a PET pattern occurred. Therefore, at least under some conditions the additional time participants spend processing misinformation may enable them to detect discrepancies and thus *reduce* the misinformation effect.

Conclusion

Our experiments provide important new evidence that initial testing sometimes reduces the misinformation effect. In the household scene paradigm, initial testing has yielded a consistent protective effect on free recall and source memory—whether the misinformation was introduced via narratives or questions (our experiments), or via social contagion (Huff et al., 2013; Huff et al., 2016). The claim that "research has consistently demonstrated that taking an initial test prior to receiving misinformation can increase misinformation susceptibility" (Gordon & Thomas, 2017, p. 190) no longer applies. Misinformation format does not dictate whether RES or a PET occurs. Likely factors to be considered include whether the witnessed event or misinformation phase presents a cohesive context, and whether the misinformation introduces additive or contradictory misinformation. Finally, despite the potential for an initial test to inoculate memory from misinformation, we found that commission errors made during that initial test are likely to persist on later tests. Thus, although taking an initial test can protect memory from misinformation, this protective factor is likely to be lost for false details that are spontaneously elicited by the initial test.

Authors' note

The authors made equivalent contributions to this project.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Mark J Huff https://orcid.org/0000-0002-0155-7877

References

Belli, R. F., Windschitl, P. D., McCarthy, T. T., & Winfrey, S. E. (1992). Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 356–367. doi:10.1037/0278-7393.18.2.356

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207–230. doi:10.1016/0010-0285(81)90008-6

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

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition*, *29*, 1120–1129. doi:10.3758/BF0306381

Chan, J. C. K., & Langley, M. 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 memory: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied*, *17*, 418–432. doi:10.1037/a0025147

Chan, J. C. K., Manley, K. D., & Lang, K. (2017). Retrieval-enhanced suggestibility: A retrospective and new investigation. *Journal of Applied Research in Memory and Cognition*, *6*, 213–229. doi:10.1016/j.jarmac.2017.07.003

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

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

Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *The Quarterly Journal of Experimental Psychology*, *70*, 1211–1235. doi:10.1080/17470219.2016.1175485

Congleton, A., & Rajaram, S. (2012). The origin of the interaction between learning method and delay in the testing effect: The roles of processing and conceptual retrieval organization. *Memory & Cognition*, *40*, 528–539. doi:10.3758/s13421-011-0168-y

Eakin, D. K., Schreiber, T. A., & Sergeant-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 813–825. doi:10.1037/02787-7393.29.5.813

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition*, *33*, 770–782. doi:10.3758/BF03193073

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavioral Research Methods, Instruments & Computers*, *28*, 1–11. doi:10.3758/BF03203630

Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a self-administered interview. *Applied Cognitive Psychology*, *26*, 568–575. doi:10.1002/acp.2828

Gordon, L. T., & Thomas, A. K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory & Cognition*, *42*, 186–197. doi:10.3758/s13421-013-0361-2

Gordon, L. T., & Thomas, A. K. (2017). The forward effects of testing on eyewitness memory: The tension between suggestibility and learning. *Journal of Memory and Language*, *95*, 190–199. doi:10.1016/j.jml.2017.01.004

Gordon, L. T., Thomas, A. K., & Bulevich, J. B. (2015). Looking for answers in all the wrong places: How testing facilitates learning of misinformation. *Journal of Memory and Language*, *83*, 140–151. doi:10.1016/j.jml.2015.03.007

Huff, M. J., Davis, S. D., & Meade, M. L. (2013). The effects of initial testing on false recall and false recognition in the social contagion of memory paradigm. *Memory & Cognition*, *41*, 820–831. doi:10.3758/s13421-013-0299-4

Huff, M. J., & Umanath, S. (2018). Evaluating the suggestibility to additive and contradictory misinformation following explicit error detection in younger and older adults. *Journal of Experimental Psychology: Applied*, *24*, 1–16. doi:10.1037/xap0000138

Huff, M. J., Weinsheimer, C. C., & Bodner, G. E. (2016). Reducing the misinformation effect though initial testing: Take two tests and recall me in the morning? *Applied Cognitive Psychology*, *30*, 61–69. doi:10.1002/acp.3167

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. doi:10.1037/0033-2909.1.3

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, *103*, 48–59. doi:10.1037/a0021977

Kay, H. (1955). Learning and retaining verbal material. *British Journal of Psychology*, *46*, 81–100. doi:10.1111/j.2044-8295.1955.tb.00527.x

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*, 85–97. doi:10.1016/j.jml.2011.04.002

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

Lane, S. M., Mather, M., Villa, D., & Morita, S. K. (2001). How events are reviewed matters: Effects of varied focus on eyewitness suggestibility. *Memory & Cognition*, *29*, 940–947. doi:10.3758/BF03195756

LaPaglia, J. A. (2013). *Testing can enhance or reduce suggestibility: The importance of contextual detail during misinformation exposure* (Doctoral dissertation). Retrieved from the Digital Repository at Iowa State University. (Paper 13263)

LaPaglia, J. A., & Chan, J. C. K. (2013). Testing increases suggestibility for narrative-based misinformation but reduces suggestibility for question-based misinformation. *Behavioral Sciences and the Law*, *31*, 593–606. doi:10.1002/bsl.2090

LaPaglia, J. A., & Chan, J. C. K. (2019). Telling a good story: The effects on memory retrieval and context processing on eyewitness suggestibility. *PLoS ONE*, *14*, e0212592. doi:10.1371/journal.pone.0212592

LaPaglia, J. A., Wilford, M. M., Rivard, J. R., Chan, J. C. K., & Fisher, R. F. (2014). Misleading suggestions can alter later memory reports even following a cognitive interview. *Applied Cognitive Psychology*, *28*, 1–9. doi:10.1002/acp.2950

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. doi:10.1037/0278-7393.16.6.1077

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*, 349–358. doi:10.3758/BF03198473

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition*, *17*, 349–358. doi:10.3758/BF03197418

Loftus, E. F. (1979). The malleability of human memory. *American Scientist*, *67*, 312–320.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into visual memory. *Journal of*

Experimental Psychology: Human Learning and Memory, *4*, 19–31. doi:10.1037/0878-7393.4.1.19

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavioral Research Methods*, *43*, 679–690. doi:10.3758/s13428-010-0049-5

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypothesis. *Journal of Experimental Psychology: General*, *114*, 1–16. doi:10.1037/0096-3445.114.1.1

Meade, M. L., & Roediger, H. L., III. (2002). Explorations in the social contagion of memory. *Memory & Cognition*, *30*, 995–1009. doi:10.3758/BF03194318

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

Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology*, *17*, 107–119. doi:10.1002/acp.857

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

Pastötter, B., & Bäuml, K.-H. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*, 286. doi:10.3389/fpsyg.2014.00286

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*, 335. doi:10.1126/science.1191465

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. doi:10.1037/a0037559

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

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 803–814. doi:10.1037/0278-7393.21.4.803

Roediger, H. L., III, Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulleting & Review*, *8*, 365–371. doi:10.3758/BF03196174

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*, 1432–1463. doi:10.1037/a0037559

Saunders, J., & Jess, A. (2010). The effects of age on remembering and knowing misinformation. *Memory*, *18*, 1–11. doi:10.1080/09658210903376268

Takarangi, M. K. T., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: The development of a new stimulus set. *Applied Cognitive Psychology*, *20*, 583–590. doi:10.1002/acp.1209

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

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

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804. doi:10.3758/BF03194105

Wilford, M. M., Chan, J. C. K., & Tuhn, S. J. (2014). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. *Journal of Experimental Psychology: Applied*, *20*, 81–93. doi:10.1037/xap0000001

Zaragoza, M. S., Belli, R. S., & Payment, K. E. (2007). Misinformation effects and the suggestibility of eyewitness memory. In M. Garry & H. Hayne (Eds.), *Do justice and let the sky fall: Elizabeth F. Loftus and her contributions to science, law, and academic freedom* (pp. 35–63). Mahwah, NJ: Lawrence Erlbaum.

Zaragoza, M. S., Lane, S. M., Ackil, J. K., & Chambers, K. L. (1997). Confusing real and suggested memories: Source monitoring and eyewitness suggestibility. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. Brainerd (Eds.), *Memory for everyday and emotional events* (pp. 401–425). Hillsdale, NJ: Lawrence Erlbaum.

Zhu, B., Chen, C., Loftus, E. F., He, Q., Chen, C., Lei, X., . . . Dong, Q. (2012). Brief exposure to misinformation can lead to long-term false memories. *Applied Cognitive Psychology*, *26*, 301–307. doi:10.1002/acp.1825