Mem Cogn (2013) 41:820–831
DOI 10.3758/s13421-013-0299-4

The effects of initial testing on false recall and false recognition in the social contagion of memory paradigm

Mark J. Huff · Sara D. Davis · Michelle L. Meade

Abstract In three experiments, participants studied photographs of common household scenes. Following study, participants completed a category-cued recall test without feedback (Exps. 1 and 3), a category-cued recall test with feedback (Exp. 2), or a filler task (no-test condition). Participants then viewed recall tests from fictitious previous participants that contained erroneous items presented either one or four times, and then completed final recall and source recognition tests. The participants in all conditions reported incorrect items during final testing (a social contagion effect), and across experiments, initial testing had no impact on false recall of erroneous items. However, on the final source-monitoring recognition test, initial testing had a protective effect against false source recognition: Participants who were initially tested with and without feedback on category-cued initial tests attributed fewer incorrect items to the original event on the final source-monitoring recognition test than did participants who were not initially tested. These data demonstrate that initial testing may protect individuals' memories from erroneous suggestions.

Keywords Recall · Recognition · False memory · Misinformation

The fallibility of human memory has received considerable attention from both the popular media and academic researchers (cf. Ayers & Reder, 1998; Gallo, 2010). Specifically, the *misinformation effect* describes a phenomenon in which individuals witness an original event and are later exposed to postevent information. When this information includes misleading details (or misinformation), witnesses frequently report having experienced the misleading details in the original event, when in fact these details never occurred (Loftus, Miller, & Burns, 1978; see Zaragoza, Belli, & Payment, 2007, for a review). This effect is particularly detrimental in situations in which high accuracy is paramount, such as eyewitness testimony. As a result, researchers have identified various procedures to reduce misinformation effects, including having participants focus on source information at test (Johnson, Hashtroudi, & Lindsay, 1993; Lindsay & Johnson, 1989; Zaragoza & Lane, 1994) and warning participants about misinformation (Chambers & Zaragoza, 2001; Echterhoff, Hirst, & Hussy, 2005; Thomas, Bulevich, & Chan, 2010). The goal of the present experiments was to explore an additional method in which an individual's memory may be protected from misleading details before they are presented—namely, that of testing.

The use of testing to improve memory performance has been well documented. The verbal-learning literature has demonstrated a consistent improvement in memory as participants completed additional tests, an effect termed *hypermnesia* (Erdelyi & Becker, 1974; Payne, 1987; Roediger & Payne, 1985). Memory researchers have also identified the *testing effect*, or the relative increase in correct memory for participants who are tested in lieu of additional study (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). Taken together, multiple-test paradigms have consistently demonstrated veridical memory benefits. Of interest to the present study is whether initial testing also benefits false memory—that is, whether completing a memory test prior to receiving misinformation about an event enhances memory for the event, and thus reduces misinformation effects (cf. Allan, Midjord, Martin, & Gabbert, 2012).

M. J. Huff (✉)
Department of Psychology, University of Calgary,
2500 University Drive NW, Calgary,
Alberta T2M 1N4, Canada
e-mail: mjhuff@ucalgary.ca

S. D. Davis · M. L. Meade
Montana State University, Bozeman, MT, USA

Some evidence supports the hypothesis that initial tests may reduce misinformation. For example, Loftus (1977) showed participants a series of slides that included a green car driving past an accident scene. Participants who were initially tested on the car color prior to receiving misinformation that the car was blue were less likely to later report the car as being blue. Furthermore, Loftus (1979) discussed an unpublished study in which participants who completed an initial free-recall test prior to the suggestion of misinformation reported fewer misinformation items on a final test than did participants who had not completed an initial free-recall test.

Nevertheless, there is also growing evidence that initial testing may *increase* the susceptibility to misinformation effects. Chan and colleagues (Chan & Langley, 2011; Chan & LaPaglia, 2011; Chan, Thomas, & Bulevich, 2009; Thomas et al., 2010) developed a paradigm in which participants viewed an event that was immediately followed by a cued-recall test (*initial-test condition*) or by a filler task (*no-initial-test condition*). Participants then listened to a narrative that contained information that was consistent with the witnessed event, as well as misleading details that had not originally been presented. On a final test, the misinformation effect was greater in the initial-test than in the no-initial-test condition. Chan and colleagues referred to heightened false memory following initial tests as a *reversed testing effect*, later termed *retrieval enhanced suggestibility* (RES). Follow-up studies have established RES as being a robust finding, presumably because the initial test calls attention to item-specific details so that participants are more receptive to later suggestions regarding those items. However, studies examining RES have used the same misinformation source (i.e., a narrative presumably prepared by the experimenter), suggesting that caution is needed when extrapolating to other false memory paradigms.

The present series of experiments extended previous research on the relation between initial testing and misinformation effects by examining the role of initial testing on false memory in the social contagion paradigm (Roediger, Meade, & Bergman, 2001). In this paradigm, participants study household scenes and then collaboratively recall the scenes with a confederate who introduces specific items that were not actually presented during study (contagion items). On subsequent individual recall and source-monitoring recognition tests, participants frequently report the contagion items that were suggested by the confederate. Of interest is whether initial testing produces RES in the social contagion paradigm (cf. Chan et al., 2009) or alternatively, protects against misinformation effects (cf. Loftus, 1977, 1979).

Like the misinformation paradigm used in previous studies of initial testing and false memory, the social contagion paradigm involves the study of an original event followed by the suggestion of misinformation. Critically, however, the social contagion paradigm also differs from the misinformation paradigm in several important ways that may further inform the parameters of initial testing effects on false memory. First, the study materials differ between the two paradigms: The misinformation paradigm typically involves presentation of a temporally ordered narrative, such as a simulated crime or a dramatic television episode. In contrast, the social contagion paradigm materials involve static images containing schematically related items. The completion of an initial test may therefore have separate effects, depending on whether memory illusions are created via schematic associations or through interference from postevent information (see Roediger & McDermott, 2000, for a discussion of memory illusions arising from different mechanisms). Specifically, we predict that initial recall testing of the schematically consistent items used in the social contagion materials will enhance organizational processing among items (cf. Zaromb & Roediger, 2010) or generate mediators that could be used as retrieval cues (Pyc & Rawson, 2010), and so improve memory for an original event, and thus protect against later errors.

Furthermore, the errant items suggested in the social contagion paradigm are additive (items that were not present in the scenes), while the errant items suggested in the misinformation paradigm are typically contradictory (items that directly oppose an originally studied item; Nemeth & Belli, 2006). Initial testing may be especially likely to influence the subsequent adoption of contradictory information because prior retrieval of a specific item should help participants detect a discrepancy between the retrieved and suggested item, and thus either reject the misinformation (Tousignant, Hall, & Loftus, 1986; see too Pansky & Tenenboim, 2011) or be more likely to incorporate the misinformation (cf. Chan et al., 2009). In contrast, additive misinformation should be less likely to elicit discrepancy detection and/or increased attention to specific items, as there are no conflicts with any one specific item from an original event. Therefore, any effect of initial testing may be confined to paradigms that involve contradictory misinformation.

In addition, the sources of the suggested information differ. Roediger et al. (2001) suggested that misinformation delivered by an experimenter (as in typical misinformation studies) may possess an informative quality that leads participants to respond with misinformation items simply because they are assumed to be true. In the social contagion paradigm, however, participants are confronted with misinformation items from a source that they deem to be as fallible as themselves—namely, a confederate or a hypothetical “previous participant” in the study. Relying on information provided from a fellow participant may reflect a different process than relying on the informative/authoritative qualities of the information delivered by the experimenter, and so may play a role in the occurrence of how initial testing influences false memory.

Finally, the social contagion paradigm includes both recall and source-monitoring recognition tests. The source-monitoring recognition test requires participants to specify where in the experiment certain items were presented: In the original event, in the postevent information, in both, or in neither. The use of a source-monitoring test may be critical to demonstrating an effect of initial testing on social contagion because the test directs attention to item-specific details that have been shown to influence testing effects on false memory (cf. Thomas et al., 2010). Furthermore, Chan and McDermott (2007; see also Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010) demonstrated that initial testing enhances recollective details, including source information, and therefore, source-monitoring recognition may be more sensitive than cued recall for determining initial testing effects. Given these paradigmatic differences between misinformation and social contagion studies, it is important to determine whether or not the initial-testing effects obtained in the misinformation paradigm will translate to the social contagion paradigm.

In addition, in the present experiments examine two novel parameters that may further inform the nature of initial testing on subsequent false memory: Feedback on the initial test, and number of repetitions. The present experiments systematically focused on the role of feedback by including an initial test without feedback (Exps. 1 and 3) and an initial test with feedback (Exp. 2). Errors produced on initial recall tests will likely show up on subsequent tests (see, e.g., Roediger, Jacoby, & McDermott, 1996), so corrective feedback on the initial test should minimize errors on subsequent tests (cf. Roediger & Marsh, 2005). Alternatively, corrective feedback may increase attention to specific items, and thus inflate errors on subsequent tests (cf. Thomas et al., 2010).

Additionally, in the present study we manipulated the number of presentations of misinformation. When false items are repeated several times, participants are likely to report these items more frequently than when false items are only presented once (Meade & Roediger, 2002). If testing has an effect on the likelihood that participants incorporate misinformation into their memory, this effect may be modulated by the number of times that the misinformation is presented, so that any effects of initial testing are reduced for information presented four times.

The first experiment reported here examines the effect of an initial category-cued recall test on subsequent false recall and recognition in the social contagion of memory paradigm. Previous research using this paradigm had established that participants incorporate misleading suggestions into their subsequent memory reports. Of interest in Experiment 1 was whether an initial category-cued recall test would influence the magnitudes of false recall and/or false recognition for misleading items suggested one or four times.

Experiment 1

Method

Participants Undergraduates at Montana State University participated for partial completion of a research requirement in an introductory psychology course. All were native English speakers with normal or corrected-to-normal vision. The data from eight of the participants were not included in the analyses because they reported suspicion or failed to follow the directions, leaving 72 participants for the analyses.

Design The experiment utilized a 2 (initial test: recall vs. arithmetic) × 3 (number of presentations: zero vs. one vs. four) × 2 (expectancy: high vs. low) mixed design with initial test as a between-subjects variable and number of presentations and expectancy as within-subjects variables.

Materials Six photographs of common household scenes, developed by Roediger et al. (2001), were used as study stimuli. These scenes depicted a toolbox, a bathroom, a kitchen, a bedroom, a closet, and a desk. The scenes were always presented in this order. The scenes contained an average of 23.8 objects and also excluded contagion items (items that would falsely be suggested by the confederate that were not actually present in the scenes). The contagion items in the original studies were updated slightly by asking 18 Montana State University undergraduate students to list items that they would expect to see when viewing the household scenes listed above. Of the items listed, the most frequently listed item for each scene that was not already presented was used as the high-expectancy contagion item. Low-expectancy items were determined by taking an item that was expected to be presented in a particular household scene but that was neither reported as frequently as the high-expectancy item nor already presented. The high- and low-expectancy contagion items for each scene can be found in the Appendix.

Fictitious recall tests were created as a method to introduce contagion information to participants (see Meade & Roediger, 2002, for a similar procedure). These recall tests were created by having five individuals write items that were correctly presented in the scenes, as well as items that were not presented in the scenes (contagion items). Each recall test contained an average of eight items in total, as pilot testing had indicated that this was the average number of total items recalled after participants had studied the household scenes. The high-expectancy contagion item was always reported in Position 4 in the fictitious recall tests, and the low-expectancy contagion item was always presented in Position 6. Correct items were randomly presented in the other positions of the recall test.

The five fictitious written recall tests were then photocopied and compiled into packets to be presented to participants as the responses of previous participants in the study. The presentation of contagion items was then manipulated within the packets of fictitious recall tests. Specifically, for two of the scenes, the contagion items were presented on zero of the five fictitious recall tests. For another two scenes, the contagion items were presented on one of the five fictitious recall tests. For the remaining two scenes, the contagion items were presented on four of the five fictitious recall tests. Thus, each packet of five writers' fictitious recall sheets contained two scenes without contagion items, two scenes with one presentation of contagion items, and two scenes with four presentations of contagion items. The contagion items were counterbalanced across the scenes and across the writers.

Procedure All participants completed the experiment individually or in groups of up to three. Participants were randomly assigned to either the initial-test condition or the no-initial-test condition. When groups were tested, all were members of the same test condition. At study, participants were presented with the household scenes one at a time for 15 s each. A title screen always preceded the household scene (e.g., "the toolbox"), and the experimenter verbally labeled each scene as it was presented. Participants were informed to pay attention to the scenes during study, as their memory would be tested later.

Following the study phase, participants in both conditions completed a filler task consisting of arithmetic problems for 4 min. Following a filler task, participants in the no-initial-test condition continued working on the arithmetic filler for an additional 12 min, while participants in the initial-test condition completed a recall test that also lasted 12 min. For each of the six scenes studied, participants were given a sheet of paper with the name of a scene printed at the top and were given 2 min to write as many items as they could remember from the scene. The scenes were tested in the same order in which they had been studied.

After the testing/filler phase, participants in both conditions were presented with the fictitious recall tests containing the contagion items. They were informed that the recall tests had been completed by five participants who had previously participated in the study. The participants were informed that an additional purpose of the experiment was to examine how the pleasantness of an item might influence memory performance, and they were asked to read each item recalled by the previous participant and to circle the items that they felt were pleasant. Participants read one recall test at a time. If items were repeated across recall tests, participants were asked to make pleasantness judgments again for the repeated items. This pleasantness manipulation was used to ensure that participants attended to each item on the recall sheets. This task was not timed, and participants took approximately 6 min to complete this task.

Immediately following the presentation of the fictitious recall tests, participants were instructed to complete a final individual recall test. This test was identical to the initial recall test, in which participants were instructed to recall items from the scenes that had originally been presented on the computer screen. Following the final recall test, participants were then given a source-monitoring recognition test containing items that had been originally presented in the studied scenes, contagion items, and items that had not been presented anywhere in the experiment. Participants were instructed to specify where each item in the test had been presented; the options included in the studied scenes, in the recall tests from the other participants, in both the scenes and the recall tests, or not in the experiment at all. Participants were given as much time as they needed to complete the source-monitoring test.

Following the source-monitoring test, participants were fully debriefed and awarded credit. In addition, participants were probed for suspicion regarding the nature of the materials. The data from participants who were suspicious of the recall tests were replaced. A typical experimental session was approximately 45 min.

Results

For all of the results reported, statistical significance was set at $p < .05$ unless otherwise noted. Effect sizes were calculated using partial eta squared (η_p^2) for analyses of variance (ANOVAs), and Cohen's d for t tests.

Correct recall The proportions of items correctly recalled on both the initial and final recall tests are reported in Table 1 for participants in the initial-test condition and in the no-initial-test condition. Analyses of the proportions of correctly recalled items demonstrated that participants in the initial-test condition significantly increased their correct recall from the initial recall test to the final recall test (.23 vs. .34), $t(35) = 12.91$, $SEM = .01$, $d = 1.56$. Furthermore, participants who completed an initial recall test reported significantly more correct items on the final recall test than did those who were not initially tested (.34 vs. .29), $t(70) = 2.65$, $SEM = .01$, $d = 0.63$, demonstrating a testing-effect advantage (Roediger & Karpicke, 2006). This finding is important, given that previous studies on initial testing and misinformation have not always obtained reliable testing effects (see Chan et al., 2009; Thomas et al., 2010). Finally, when comparing across test conditions, correct recall was greater on the recall test in the no-initial-test condition than on the initial test in the initial-test condition (.29 vs. .23), $t(70) = 3.05$, $SEM = .01$, $d = 0.73$, most likely due to relearning during the contagion phase.

False recall Proportions of contagion items falsely recalled on the final recall test are reported in Table 2. A 2 (initial test) × 3 (presentation: zero vs. one vs. four) × 2 (expectancy: high vs. low) mixed-factorial ANOVA computed on the proportions of falsely recalled contagion items revealed a main effect of presentation, $F(2, 140) = 46.04$, $MSE = .10$, $\eta_p^2 = .40$. Post hoc *t* tests confirmed significant social contagion effects for both one and four presentations: Contagion items presented one time were recalled more frequently than contagion items presented zero times (i.e., not presented; .15 vs. .06), $t(71) = 3.31$, $SEM = .03$, $d = 0.59$, and contagion items presented four times were falsely recalled more frequently than contagion items presented zero times (.40 vs. .06), $t(71) = 9.92$, $SEM = .03$, $d = 1.43$, and than contagion items presented only once (.40 vs. .15), $t(71) = 5.42$, $SEM = .05$, $d = 1.00$. Critically, however, the main effect of initial test failed to reach significance, $F < 1$, $\eta_p^2 = .01$, demonstrating that testing neither increased recall of contagion items (an RES effect) nor decreased recall of contagion items (a protective effect of testing). The nonsignificant comparisons that we report were each further tested using a Bayesian estimate of the strength of evidence supporting the null hypothesis, the Bayesian information criterion (BIC; Masson, 2011; Wagenmakers, 2007). In this analysis, a model that assumes an effect is contrasted with a model that assumes no effect. When comparing the initial-test and no-test conditions, the estimated probability that the null-effect model was preferred over a model assuming a testing difference was $p_{BIC} = .89$, thus providing evidence favoring the null hypothesis.

However, a significant Initial Test × Expectancy interaction emerged, $F(1, 140) = 11.32$, $MSE = .06$, $\eta_p^2 = .32$. Follow-up *t* tests indicated that false recall of high-expectancy items was marginally higher in the initial-test than in the no-initial-test condition (.28 vs. .20), $t(70) = 1.93$, $SEM = .06$ $p = .06$, $d = 0.46$, $p_{BIC} = .57$, while false recall of low-expectancy items was lower in the initial-test than in the no-initial-test condition (.13 vs. .20), $t(70) = 1.97$, $SEM = .03$, $p = .05$, $d = 0.49$. The Initial Test × Expectancy interaction was not predicted, and we suspect that the interaction is driven primarily by the high baseline error rate for the high-expectancy items in the initial-test condition (we addressed and corrected for this issue is Exp. 2).

The main effects of presentation and expectancy were also qualified by a significant Presentation × Expectancy interaction, $F(1, 140) = 3.10$, $MSE = .06$, $\eta_p^2 = .04$. Follow-up *t* tests revealed that, when contagion items were not suggested (i.e., scenes with zero contagion presentations), high-expectancy items were recalled more frequently than low-expectancy items (.09 vs. .03), $t(71) = 2.11$, $SEM = .03$, $d = 0.33$, demonstrating that high-expectancy contagion items were intruded more frequently when they were not suggested. This pattern was also found for one-presentation contagion items: High-expectancy items were recalled more frequently than low-expectancy items (.22 vs. .07), $t(71) = 3.68$, $SEM = .04$, $d = 0.61$. Interestingly, expectancy did not influence the contagion effect for items presented four times (.40 vs. .39, for high and low expectancy, respectively), $t < 1$, $d = 0.03$, $p_{BIC} = .92$. The Presentation × Initial Test interaction, as well as the three-way interaction, failed to reach significance, both $F\text{s} < 1$, $\eta_p^2\text{s} < .01$, $p_{BIC}\text{s} > .98$.

Source-monitoring recognition Performance on the source-monitoring recognition test is displayed on Table 3, which shows proportions of source attributions for contagion items as a function of the number of presentations, test condition, and the four response types (presented in the studied scene, presented in the scene and recalled by another participant, only recalled by another participant, or neither presented in the scene nor recalled by other participants). Total false recognition was operationalized as the proportion of contagion items that participants recognized as having been presented in the scenes ("scene only" responses plus "scene and other participant" responses). Responding to contagion items using one of these source responses indicated false recognition that the contagion items had been presented in the studied scenes (see Meade & Roediger, 2002, for an identical scoring procedure).

To examine differences in total false recognition, a 2 (initial test) × 3 (presentation) mixed-factorial ANOVA was used. A main effect of presentation was obtained, $F(2, 140) = 27.63$, $MSE = .06$, $\eta_p^2 = .28$, and follow-up tests demonstrated that presenting contagion items four times increased false recognition relative to when contagion items were not presented (.59 vs. .28), $t(71) = 7.17$, $SEM = .04$, $d = 1.08$, or were presented once (.59 vs. .41), $t(71) = 4.21$, $SEM = .04$, $d = 0.56$. Additionally, presenting contagion items only once also inflated false recognition for the

Table 1 Mean proportions of correct recall on initial and subsequent recall tests as a function of initial test for Experiments 1 (initial test without feedback), 2 (initial test with feedback), and 3 (initial test without feedback)

	Initial Test		No Initial Test
	Test 1	Test 2	
Experiment 1	.23 (.06)	.34 (.08)	.29 (.08)
Experiment 2	.21 (.05)	.32 (.06)	.28 (.07)
Experiment 3	.23 (.05)		

Standard deviations are listed in parentheses. The recall test in the no-initial-test condition occurred at the same time in Experiments 1 and 2 as did Test 2 in the initial-test condition. No final test was completed for Experiment 3. Therefore, in that experiment the proportion of correct recall is reported only for the initial test in the initial-test condition

Table 2 Mean proportions of false recall as a function of initial test, number of presentations, and expectancy for Experiments 1 (initial test without feedback) and 2 (initial test with feedback)

	Initial Test		No Initial Test	
	High	Low	High	Low
Experiment 1				
Presentation condition				
Zero	.14 (.28)	.00 (.00)	.04 (.14)	.06 (.16)
One	.22 (.25)	.03 (.12)	.22 (.30)	.11 (.27)
Four	.47 (.36)	.36 (.33)	.33 (.41)	.42 (.37)
Experiment 2				
Presentation condition				
Zero	.07 (.18)	.03 (.12)	.01 (.08)	.01 (.08)
One	.07 (.18)	.14 (.23)	.19 (.30)	.08 (.19)
Four	.28 (.30)	.19 (.27)	.38 (.37)	.24 (.30)

Standard deviations are listed in parentheses. Means in the initial-test condition reflect performance on Test 2. The zero-presentation condition reflects performance on those scenes in which contagion items were not presented

contagion items relative to when they were not presented (.41 vs. .28), *t*(71) = 3.33, *SEM* = .04, *d* = 0.45, thereby replicating the same pattern of social contagion effects found in false recall.

Most importantly, a main effect of initial test emerged, *F* (1, 70) = 5.02, *MSE* = .13, $\eta_p^2 = .07$, demonstrating that participants who had completed an initial recall test were *less* likely to falsely recognize suggested contagion items as having been presented in a studied scene than were those who had not (.37 vs. .48). This pattern demonstrates that the completion of an initial recall test can reduce the probability that participants will report suggested information on a subsequent test when attention is directed to source information. Furthermore, testing differences did not vary as a function of the number of presentations, as the Presentation × Initial Test interaction failed to reach significance, $F < 1$, $\eta_p^2 < .01$, $p_{\mathrm{BIC}} = .98$.

The proportion of correct contagion recognition, or the proportion of recognition responses that were correctly attributed to the other participants and not to the scenes (the “other participants only” category) were analyzed using a 2 (initial test) × 3 (presentation) mixed-factorial ANOVA.

A main effect of presentation was found, *F*(2, 140) = 21.35, *MSE* = .06, $\eta_p^2 = .23$. Follow-up tests revealed that contagion items presented four times were more likely to be correctly attributed to the other participants than were contagion items that had not been presented (.31 vs. .11), *t*(71) = 4.72, *SEM* = .04, *d* = 0.82, but were not more likely to be correctly recognized than those presented once (.31 vs. .35), $t < 1$,

Table 3 Mean proportions of false and veridical source attributions for contagion items as a function of initial test and number of presentations for Experiments 1 and 3 (initial test without feedback) and Experiment 2 (initial test with feedback)

	Initial Test			No Initial Test		
Presentations	Zero	One	Four	Zero	One	Four
Experiment 1						
Scene only	.17 (.21)	.16 (.29)	.19 (.31)	.26 (.24)	.29 (.24)	.21 (.32)
Scene and other participant	.07 (.17)	.17 (.27)	.34 (.33)	.06 (.12)	.19 (.26)	.43 (.40)
Total false recognition	**.24 (.26)**	**.33 (.33)**	**.54 (.33)**	**.32 (.24)**	**.48 (.28)**	**.64 (.30)**
Other participants only	.08 (.14)	.35 (.29)	.36 (.32)	.14 (.20)	.34 (.29)	.26 (.28)
Neither slide nor participants	.68 (.30)	.31 (.34)	.10 (.21)	.54 (.27)	.17 (.19)	.09 (.19)
Experiment 2						
Scene only	.12 (.18)	.12 (.18)	.09 (.21)	.14 (.23)	.16 (.23)	.05 (.15)
Scene and other participant	.09 (.19)	.19 (.24)	.43 (.33)	.14 (.19)	.28 (.29)	.60 (.34)
Total false recognition	**.21 (.26)**	**.30 (.27)**	**.52 (.33)**	**.28 (.29)**	**.43 (.34)**	**.64 (.33)**
Other participants only	.16 (.22)	.41 (.23)	.44 (.34)	.13 (.19)	.34 (.27)	.30 (.31)
Neither slide nor participants	.63 (.33)	.29 (.23)	.03 (.11)	.58 (.34)	.23 (.24)	.05 (.12)
Experiment 3						
Scene only	.10 (.15)	.05 (.13)	.07 (.16)	.17 (.28)	.13 (.18)	.06 (.17)
Scene and other participant	.13 (.21)	.22 (.27)	.35 (.32)	.17 (.21)	.31 (.27)	.60 (.30)
Total false recognition	**.24 (.26)**	**.27 (.28)**	**.42 (.36)**	**.33 (.31)**	**.44 (.30)**	**.67 (.28)**
Other participants only	.21 (.25)	.46 (.28)	.45 (.34)	.09 (.15)	.32 (.28)	.27 (.29)
Neither slide nor participants	.57 (.28)	.27 (.22)	.13 (.21)	.58 (.34)	.24 (.21)	.06 (.16)

Standard deviations are listed in parentheses. Total false recognition is equal to the sum of the scene-only and scene-and-other-participant means (in bold). The zero-presentation conditions reflect performance on those scenes in which contagion items were not presented

$d = 0.14$, $p_{\mathrm{BIC}} = .90$. Additionally, presenting contagion items once increased correct recognition relative to when contagion items were not presented (.35 vs. .11), $t(71) = 6.67$, $SEM = .03$, $d = 1.01$. Considered together, increasing the number of presentations reliably increased incorrect source attributions of contagion items to the scenes but did not increase correct source attributions of contagion items to the other participants only.

The main effect of initial test failed to reach significance, $F < 1$, $\eta_p^2 < .01$, $p_{\mathrm{BIC}} = .96$, as did the Presentation × Initial Test interaction, $F(2, 140) = 1.93$, $MSE = .06$, $p = .15$, $\eta_p^2 = .03$, $p_{\mathrm{BIC}} = .96$. Thus, the completion of an initial recall test did not appear to increase the ability of participants to correctly attribute erroneously suggested contagion items to the fictitious recall sheets.

Finally, an independent-samples *t* test revealed that veridical source recognition (i.e., the proportion of studied items correctly attributed to the scene only or to both the scene and the other participant) was equivalent between the initial-test and no-initial-test conditions (.58 vs. .58), $t < 1$, $d = .01$, $p_{\mathrm{BIC}} = .92$.

Discussion

The primary findings from Experiment 1 indicated that participants in both the initial-test condition and the no-initial-test condition reported the same numbers of contagion items on the final recall test, demonstrating neither an RES effect nor a protective effect of testing. Furthermore, two secondary findings showed that increasing the number of presentations subsequently increased false recall of contagion items and that high-expectancy items were falsely recalled more frequently than low-expectancy items. Importantly, when attention was drawn to the source of contagion items on the source-monitoring recognition test, participants who were initially tested were less likely to report suggested contagion items as having been present in the scenes.

One possible explanation for why initial testing did not influence recall is that the participants who were tested initially were already reporting the contagion items (because of the schematic nature of the scenes; cf. Brewer & Treyens, 1981) and that these contagion items were persisting on subsequent tests (cf. Bartlett, 1932; Kang et al., 2011; Kay, 1955; Lane, Mather, Villa, & Morita, 2001; McDermott, 1996; Roediger et al., 1996). An examination of the proportions of contagion items reported on the initial recall test demonstrated that participants reported 9 % of the contagion items prior to their suggestion, a rate that was significantly greater than zero, $t(35) = 6.44$, $SEM = .01$. Furthermore, recalling a given contagion item on the initial recall test was highly correlated with recalling the same contagion item on the final recall test, $r = .84$, $p < .001$, regardless of whether this item was suggested or not. Considered together, it is possible that the lack of testing differences in recall may be due to the persistence of contagion items from the initial to the final recall test.

The purpose of Experiment 2 was to further examine the nature of the effect of initial testing on false recall, by including corrective feedback on the initial recall test. Past research has demonstrated that when participants are given feedback on memory errors, they are likely to eliminate these errors on subsequent tests (Pashler, Cepeda, Wixted, & Rohrer, 2005). In fact, Butler, Karpicke, and Roediger (2007) reported that when participants were given immediate feedback on a multiple-choice test, they corrected about half of their initially incorrect responses, as compared to 11 % of their incorrect responses when they were not given corrective feedback (although see Kanter & Lindsay, 2010, for an exception). Therefore, Experiment 2 was designed to replicate Experiment 1, with one critical exception: The participants in Experiment 2 were given corrective feedback on their initial recall tests. If uncorrected errors produced on the initial test in Experiment 1 inflated the errors produced on the final recall test, corrective feedback on the initial test in Experiment 2 should minimize those persisting errors, and thus offer a more complete picture of how initial testing affects subsequent false recall. Furthermore, if testing benefits on the source test were related to the attention drawn to the sources of items during recognition (as is suggested in the RES literature), feedback should further increase attention to items produced on the initial test, and so may exaggerate any effects of initial testing. We again included the final source test (as in Exp. 1) to gain further confidence about possible differences in any protective nature of initial testing across recall and recognition tests.

Experiment 2

Method

Participants Undergraduates at the University of Calgary participated for partial completion of a research requirement. All were fluent English speakers with normal or corrected-to-normal vision. Of these participants, ten were not included in the analysis due to reporting prior knowledge of the misinformation effect or suspicion of misinformation from the experimenter; an additional five were not included due to failure to follow directions, and one due to experimenter error. Therefore, the data from 72 participants were used in the analyses. All participants were tested individually.

Design The experiment utilized a 2 (initial test: recall with feedback vs. arithmetic) × 3 (number of presentations: zero vs. one vs. four) × 2 (expectancy: high vs. low) mixed-subjects design, with initial test as a between-subjects variable.

Materials and procedure The same stimulus materials and procedure as in Experiment 1 were used in Experiment 2, with the following exceptions. First, for participants in the initial-test condition, all items were marked by the experimenter as correct or incorrect on each recall sheet completed by the participant. Additionally, participants were given 20 s to examine the feedback for the recall sheet from each scene. Participants were also instructed to place a check mark next to each item to ensure that they had reexamined each item and that they had evaluated the feedback. For participants in the no-initial-test condition, the time given to complete the arithmetic filler was increased to 19 min, to equate for the time required for the initial-test condition to complete the initial test and evaluate the feedback provided by the experimenter.

Results and discussion

Effectiveness of feedback The feedback appeared to reduce the contagion items on the second test if they were recalled on the first test. Specifically, we found a strong negative correlation between contagion items recalled on the first test and the same contagion items recalled on the second test, $r = -.674$, $p < .001$. Furthermore, this correlation was found despite the later presentation of contagion items.

Correct recall The proportions of items correctly recalled on the initial and final recall tests completed by participants in the initial-test and no-initial-test conditions are reported in Table 1. Replicating Experiment 1, participants in the initial-test condition reliably increased their correct recall from the initial to the final recall test (.21 vs. .32), $t(35) = 14.02$, $SEM = .01$, $d = 1.99$. Further, replicating Experiment 1, a significant testing effect was found: Participants who completed an initial recall test had significantly greater correct recall on the final recall test, relative to the participants in the no-initial-test condition (.32 vs. .28), $t(70) = 2.58$, $SEM = .01$, $d = 0.67$. Similarly, across test conditions, final correct recall was greater in the no-initial-test condition than on the initial test when it was given (.28 vs. .21), $t(70) = 5.15$, $SEM = .01$, $d = 1.23$, most likely due to relearning during the contagion phase.

False recall The proportions of contagion items falsely recalled on the final recall test are reported in Table 2. A 2 (initial test) × 3 (presentation: zero vs. one vs. four) × 2 (expectancy: high vs. low) mixed-factorial ANOVA was computed on the proportions of falsely recalled contagion items. A main effect of presentation was found, $F(2, 140) = 33.42$, $MSE = .06$, $\eta_p^2 = .32$. Replicating Experiment 1, false recall was greater for contagion items presented one time than for contagion items presented zero times (i.e., not presented; .12 vs. .03), $t(71) = 3.63$, $SEM = .02$, $d = 0.66$. False recall was also greater for contagion items presented four times than for contagion items presented zero times (.27 vs. .03), $t(71) = 8.38$, $SEM = .03$, $d = 1.43$, or presented once (.27 vs. .12), $t(71) = 4.29$, $SEM = .03$, $d = 0.76$.

Importantly, and consistent with Experiment 1, the main effect of initial test failed to reach significance, $F(1, 70) = 1.32$, $MSE = .06$, $p = .26$, $\eta_p^2 = .02$, $p_{\mathrm{BIC}} = .81$. That is, participants who completed an initial test with feedback were neither more nor less likely than participants in the no-test condition to incorporate misleading items into their subsequent recall. Also consistent with Experiment 1, a main effect of expectancy was found, $F(1, 140) = 5.84$, $MSE = .06$, $\eta_p^2 = .08$, demonstrating that high-expectancy items were more likely to be falsely recalled than low-expectancy items (.17 vs. .12). All interactions, including the Initial Test × Expectancy interaction and the three-way interaction, failed to reach significance, all Fs < 2.37, all ps > .13, η_p^2s < .04, p_{BIC}s > .96.

Source-monitoring recognition Responses on the source-monitoring recognition test are presented in Table 3. The proportions of total false recognition were analyzed using a 2 (initial test) × 3 (presentation) mixed-factorial ANOVA. A significant main effect of presentation was found $F(2, 140) = 29.38$, $MSE = .07$, $\eta_p^2 = .30$, demonstrating that the social contagion effect persisted on the final source test. Follow-up tests confirmed greater false recognition for contagion items with four presentations than for those with zero presentations (.58 vs. .25), $t(71) = 7.79$, $SEM = .04$, $d = 1.06$, or one presentation (.58 vs. .37), $t(71) = 4.58$, $SEM = .05$, $d = 0.65$. False recognition was also greater in scenes with one than with zero presentations (.37 vs. .25), $t(71) = 2.84$, $SEM = .04$, $d = 0.41$.

Critically, a main effect of initial test was also found, $F(1, 70) = 4.56$, $MSE = .14$, $\eta_p^2 = .07$. Replicating Experiment 1, participants who completed an initial test (this time with feedback) demonstrated reduced false source recognition relative to participants in the no-initial-test condition (.34 vs. .45). The interaction between initial test and the number of presentations failed to reach significance, $F < 1$, $\eta_p^2 < .01$, $p_{\mathrm{BIC}} = .98$.

The proportions of correct contagion recognition were analyzed using a 2 (initial test) × 3 (presentation) mixed-factorial ANOVA. Again, a contagion attribution was considered correct when participants correctly attributed contagion items to the fictitious recall sheets. A main effect of presentation was found, $F(2, 140) = 20.77$, $MSE = .06$, $\eta_p^2 = .23$. Follow-up tests completed on correct contagion attributions revealed that contagion items presented four times were more likely to be correctly attributed to the fictitious recall sheets than were those with zero presentations (.37 vs. .15), $t(71) = 5.08$, $SEM = .05$, $d = 0.95$. Furthermore, correct attributions were also more likely for contagion items presented once rather than zero times (.38 vs. .15), $t(71) = 7.03$, $SEM = .03$, $d = 0.83$. Four presentations did not increase correct attributions relative to one presentation (.37 vs. .38), $t < 1$, $d = 0.03$, $p_{\mathrm{BIC}} = .97$.

Interestingly, a main effect of initial test was also found, $F(1, 70) = 3.96$, $MSE = .09$, $p = .05$, $\eta_p^2 = .05$, which demonstrates that across presentations, initial testing with feedback improved correct source attributions relative to participants who were not tested initially (.33 vs. .25). Note that this finding differs from that in Experiment 1: Initial testing with feedback in Experiment 2 served both to reduce misattributions and to increase correct source attributions for contagion items. The interaction between initial testing and number of presentations failed to reach significance, $F(2, 140) = 1.05$, $MSE = .06$, $p = .35$, $\eta_p^2 = .02$, $p_{\mathrm{BIC}} = .89$.

Finally, veridical source recognition of filler items did not differ between the initial-test and no-initial-test conditions (.53 vs. .56), $t < 1$, $d = 0.15$, $p_{\mathrm{BIC}} = .88$.

In sum, Experiment 2 demonstrated that either completing an initial test with feedback or not completing an initial test produced the same rates of false recall of contagion items on a final test. Similar to Experiment 1, testing with feedback produced neither an RES pattern nor a protective effect of testing on recall. Critically, however, testing with feedback did influence false recognition on a source-monitoring test by reducing the probability at which participants attributed the contagion items to the original scenes. This pattern also occurred regardless of the number of times that participants were exposed to contagion items.

It is important to note, however, that the source-monitoring tests in Experiments 1 and 2 were always preceded by a final recall test. The conclusion that testing protects against false recognition would be strengthened by demonstrating that the protective effect persists on recognition tests not confounded by prior recall (e.g., Gallo & Roediger, 2002; Roediger et al., 1996). Given that the first two experiments revealed no differences on recall tests between the initial-test and no-initial-test conditions, we hypothesized that the protective effect of testing would also occur on recognition tests not confounded by prior recall.

Experiment 3

Method

Participants A group of 72 undergraduates from the University of Calgary participated for partial completion of a research requirement. All were fluent English speakers with normal or corrected-to-normal vision.

Design The experiment utilized a 2 (initial test: recall vs. arithmetic) × 3 (number of presentations: zero vs. one vs. four) mixed-subjects design, with initial test manipulated between subjects.

Materials and procedure The same materials and procedure were used as in Experiment 1, with one exception: Participants in both the initial-test and no-initial-test conditions did not complete a final recall test. Instead, both conditions immediately completed the same source-monitoring test used in Experiment 1. Feedback was also eliminated in Experiment 3, as the protective effect in source-monitoring recognition had occurred in both the presence and absence of feedback.

Results and discussion

Initial-test correct recall The proportion of correctly recalled items on the recall test in the initial-test condition is reported in Table 1. On average, participants correctly recalled 23 % of the items in the scenes.

Source-monitoring recognition Responses on the source-monitoring recognition test are presented in Table 3. The proportions of total false recognition were analyzed using a 2 (initial test) × 3 (presentation) mixed-factorial ANOVA. A main effect of presentation was found, $F(2, 140) = 19.60$, $MSE = .07$, $\eta_p^2 = .22$, demonstrating greater false recognition with four presentations of contagion items than with either zero presentations (.54 vs. .28), $t(71) = 5.54$, $SEM = .05$, $d = 0.85$, or one presentation (.54 vs. .35), $t(71) = 4.18$, $SEM = .05$, $d = 0.59$. False recognition was marginally greater in scenes with one presentation than when contagion items were not presented (.35 vs. .28), $t(71) = 1.91$, $SEM = .03$, $p = .06$, $d = 0.25$, $p_{\mathrm{BIC}} = .81$.

Importantly, a main effect of initial test was also found, $F(1, 70) = 12.59$, $MSE = .13$, $\eta_p^2 = .15$. Replicating Experiments 1 and 2, completing an initial test reduced false source recognition relative to no testing (.31 vs. .48). The Initial Test × Presentation interaction was not significant, $F(2, 140) = 1.62$, $MSE = .11$, $p = .20$, $\eta_p^2 = .02$, $p_{\mathrm{BIC}} = .97$. Therefore, initial testing protected participants from false recognition even when the recognition test was not confounded by prior recall (as in Exps. 1 and 2).

The proportions of correct contagion recognition were also analyzed using a 2 (initial test) × 3 (presentation) mixed-factorial ANOVA. A main effect of presentation was found $F(2, 140) = 19.80$, $MSE = .06$, $\eta_p^2 = .22$, demonstrating that presenting contagion items four times increased correct source attributions to the fictitious recall sheets more than when contagion items were not presented (.36 vs. .15), $t(71) = 4.41$, $SEM = .05$, $d = 0.76$. Correct attributions were also more likely following one rather than zero presentations (.39 vs. .15), $t(71) = 6.79$, $SEM = .04$, $d =$

0.95. As in Experiment 2, however, correct attributions did not differ between four presentations and one presentation (.36 vs. .39), $t < 1$, $d = 0.10$, $p_{\mathrm{BIC}} = .91$.

Also consistent with Experiment 2, a main effect of initial test showed that correct contagion recognition was greater in the initial-test condition than in the no-test condition (.37 vs. .23), $F(1, 70) = 11.96$, $MSE = .10$, $\eta_p^2 = .15$. The Presentation × Initial Test interaction was not significant, $F < 1$, $\eta_p^2 < .01$, $p_{\mathrm{BIC}} = .99$.

Finally, veridical recognition was greater following the no-initial-test condition than following the initial-test condition (.66 vs. .46), $t(71) = 4.36$, $SEM = .03$, $d = 1.03$. This pattern differs from Experiments 1 and 2, and suggests that a final recall test influences the recognition of correctly studied items.

General discussion

Across three experiments, our primary objective was to examine the influence of initial testing on the adoption of incorrect socially suggested items on subsequent recall and source recognition tests in the social contagion of memory paradigm. Importantly, these experiments demonstrated that completing an initial test can reduce the susceptibility to contagion items, but that this reduction is contingent on the type of final test completed. In Experiments 1 and 2, completing an initial test had no impact on final false recall of misleading contagion items. However, across all three experiments, completing an initial test had a protective effect against false recognition on a source-monitoring test. The protective effect of initial testing occurred both when the recognition test was confounded by prior recall (Exps. 1 and 2) and not confounded by prior recall (Exp. 3).

Null effects of initial testing on recall occurred when the initial test was taken either without feedback (Exp. 1) or with feedback (Exp. 2). In both experiments, participants demonstrated reliable social contagion effects that were greater when misleading contagion items were presented four times relative to one time, and the null effect of initial testing was consistent across one and four presentations. The lack of initial-testing effects on false recall in the social contagion paradigm is inconsistent with previously reported evidence that initial testing both protects against misinformation (Loftus, 1977, 1979) and increases susceptibility to misinformation (Chan et al., 2009; Thomas et al., 2010). The results of the present study support neither of these patterns, suggesting that the initial testing effects obtained in the misinformation paradigm do not translate to the social contagion paradigm.

Several methodological differences between the misinformation paradigm and the social contagion paradigm may explain these discrepant results. In the typical misinformation paradigm, an experimenter introduces contradictory misinformation about a narrative event. In the social contagion paradigm, a fellow participant introduces additive misinformation about a schematic scene. In addition, the type of *initial* test differs. Misinformation studies typically rely on an item-based cued-recall test in which participants are required to report a specific item from the original narrative. In contrast, our category-cued recall test had participants freely recall items from the study scenes without additional test parameters. We may have failed to find an effect of initial testing on false recall in the social contagion paradigm because the initial test was not item-specific (cf. Hamaker, 1986; Pansky & Tenenboim, 2011; Thomas et al., 2010). Therefore, the source of misinformation (experimenter vs. fictitious participant), the study materials (narrative vs. schematic scenes), the type of suggested items (contradictory vs. additive), and/or the type of initial test (item-specific vs. non-item-specific) may be important factors that influence the role of initial testing on false recall.

Importantly, the present study demonstrated that initial testing did protect against misleading suggestions on a final source-monitoring recognition test. Across all three experiments, participants who completed an initial test were *less* likely to falsely attribute the contagion items to the scenes. Again, this protective effect of initial testing occurred when the initial test was taken with or without corrective feedback and regardless of whether or not the recognition test was confounded by prior recall. Importantly, the protective effect did not interact with the number of presentations, so that protective effects were obtained with both relatively higher and relatively lower levels of overall false recognition. The finding that testing reduces false memory on a source-monitoring test is at odds with Chan, Wilford, and Hughes (2012), who demonstrated that, in the misinformation paradigm, RES is obtained on some source tests. However, the general finding in the present study that the effect of initial testing varied across recall and recognition is consistent with broader findings in the literature that testing effects differ across recall and recognition tests (see Chan & McDermott, 2007, for a review).

A possible explanation for the protective effects of initial testing on source recognition is that initial testing may increase the recollection of details, particularly source details, of the original scenes. That is, testing immediately after study improves the memory for the specific items and source details that accompany those items. As was demonstrated by Chan and McDermott (2007), recollection can benefit subsequent recognition, especially when the recognition test contains semantically related lures (as were used in the present study), because recollection is required for discriminating between old and new related items. Importantly, Chan and McDermott also showed that initial testing may enhance recollective details, even in cases in which it does not influence overall hit rates. In the present study, initial testing did not influence correct recognition in Experiments 1 and 2, and it reduced correct recognition in Experiment 3, but it also reduced false recognition across all three experiments.

We suggest that the protective effect of initial testing against false recognition operated selectively on the source recognition test because the test instructions directed attention to the source of the test items. Therefore, the strengthening of recollective source details due to the initial test gave these participants an advantage on the source-monitoring recognition test. That is, participants may have been directed to use recollective details (i.e., where in the experiment the items had been encountered) on the source-monitoring test to reduce false recognition. In contrast, on the category-cued recall test, participants may not have been directed to use recollective details. Initial testing selectively influenced false recognition in the social contagion paradigm because the recognition test required the use of recollective source details.

The present experiments provide evidence that initial testing enhances recollective details. Across experiments, completing an initial test reduced incorrect source attributions of contagion items, and further, in Experiments 2 and 3, initial testing also increased correct source attributions of contagion items. Thus, it appears that initial testing, regardless of feedback or the presence of a final recall test, decreases incorrect source attributions, but feedback and/or final recall may play a larger role in increasing correct source attributions.

In sum, we suggest that the role of initial testing in the susceptibility to misinformation in the social contagion paradigm depends on the type of test administered. Across experiments, initial testing had no impact on false recall but consistently reduced false recognition on a final source-monitoring recognition test, a novel finding. The result that the completion of an initial test can reduce recognition errors in some paradigms is encouraging and suggests a potentially important technique for protecting against false memories.

Author note Thanks to Tabitha Knight, Chase Ladd, and Vladimir Perga for their assistance running participants. Special thanks to Daniel Gigone and Glen Bodner for helpful comments while developing this project.

Appendix

Table 4 High- and low-expectancy contagion items for the six study scenes

	Expectancy	
Scene	High	Low
Toolbox	Screws	Ruler
Bathroom	Toothbrush	Hairbrush
Kitchen	Toaster	Napkins
Bedroom	Backpack	Night stand
Closet	Shoes	Belt
Desk	Printer	Rolodex

References

Allan, K., Midjord, J. P., Martin, D., & Gabbert, F. (2012). Memory conformity and the perceived accuracy of self versus other. *Memory and Cognition, 40,* 280–286. doi:10.3758/s13421-011-0141-9

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin and Review, 5,* 1–21.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

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. doi:10.1080/09658211003702163

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

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 4,* 273–281.

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

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, 418–432.* doi:10.1037/a0025147

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. doi:10.1037/0279-7393.33.2.431

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. W., & Hughes, K. L. (2012). Retrieval can increase or decrease suggestibility depending on how the memory is tested: The importance of source complexity. *Journal of Memory and Language.* doi:10.1016/j.jml.2012.02.006

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

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not for words in multiple recall trials. *Cognitive Psychology, 6,* 159–171. doi:10.1016/0010-0285(74)90008-5

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory and Cognition, 38,* 833–848. doi:10.3758/MC.38.7.833

Gallo, D. A., & Roediger, H. L., III. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47,* 469–497.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56,* 212–242.

Johnson, M. D., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. doi:10.1037/0033-2909.114.1.3

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

Kanter, J., & Lindsay, D. S. (2010). Can corrective feedback improve recognition memory? *Memory and Cognition, 38,* 389–406.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968. doi:10.1126/science.1152408

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

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

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

Loftus, E. F. (1977). Shifting human color memory. *Memory and Cognition, 5,* 696–699.

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 a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19–31.

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

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35,* 212–230. doi:10.1006/jmla.1996.0012

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

Nemeth, R. J., & Belli, R. F. (2006). The influence of schematic knowledge on contradictory versus additive misinformation: False memory for typical and atypical items. *Applied Cognition Psychology, 20,* 563–573.

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

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning Memory and Cognition, 31,* 3–8. doi:10.1037/0278-7393.31.1.3

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101,* 5–27.

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

Roediger, H. L., III, Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language, 35,* 300–318. doi:10.1006/jmla.1996.0017

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: 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, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning Memory and Cognition, 31,* 1155–1159. doi:10.1037/0278-7393.31.5.1155

Roediger, H. L., III, & McDermott, K. B. (2000). Distortions of memory. In E. Tulving & F. Craik (Eds.), *The Oxford handbook of memory* (pp. 149–162). New York: Oxford University Press.

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

Roediger, H. L., III, & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory and Cognition, 13,* 1–7.

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 and Cognition, 14,* 329–338. doi:10.3758/BF03202511

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

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. 5–64). Mahwah: Erlbaum.

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

Zaromb, F. M., & Roediger, H. L., III. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory and Cognition, 38,* 995–1008. doi:10.3758/MC.38.8.995