Reducing the Misinformation Effect Through Initial Testing: Take Two Tests and Recall Me in the Morning?

MARK J. HUFF[1]*, CAMILLE C. WEINSHEIMER[2] and GLEN E. BODNER[2]

[1]*Department of Psychology, Washington University in St. Louis, St. Louis, USA*
[2]*Department of Psychology, University of Calgary, Calgary, Alberta, Canada*

Summary: Initial retrieval of an event can reduce people's susceptibility to misinformation. We explored whether protective effects of initial testing could be obtained on final free recall and source-monitoring tests. After studying six household scenes (e.g., a bathroom), participants attempted to recall items from the scenes zero, one, or two times. Immediately or after a 48-hour delay, non-presented items (e.g., soap and toothbrush) were exposed zero, one, or four times through a social contagion manipulation in which participants reviewed sets of recall tests ostensibly provided by other participants. A protective effect of testing emerged on a final free recall test following the delay and on a final source-memory test regardless of delay. Taking two initial tests did not increase these protective effects. Determining whether initial testing will have protective (versus harmful) effects on memory has important practical implications for interviewing eyewitnesses. © 2015 The Authors. Applied Cognitive Psychology published by John Wiley & Sons, Ltd.

Researchers have long sought to discover effective methods for improving memory accuracy. Techniques such as distinctive processing can enhance encoding (e.g., Huff, Bodner, & Fawcett, 2015; Hunt & Worthen, 2006), while warnings or penalties for errors can enhance retrieval by increasing memory monitoring (e.g., Gallo, Roediger, & McDermott, 2001; Chambers & Zaragoza, 2001). Taking an initial memory test can improve memory by facilitating encoding and/or retrieval processes. Initial testing provides retrieval practice, which can yield robust memory benefits (Roediger & Karpicke, 2006; for a review, see Rawson & Dunlosky, 2011). The present article examined whether such retrieval practice can enhance memory accuracy in a social contagion misinformation paradigm that elicits high rates of memory errors.

In the misinformation paradigm, participants are exposed to misleading details about a previous event. On a final test, misleading details are reported or endorsed more frequently relative to when those details had not been exposed to participants (e.g., Loftus, Miller, & Burns, 1978; Zaragoza, Belli, & Payment, 2007). By extension, eyewitnesses exposed to misleading details are also likely to unwittingly incorporate misinformation into their testimony. To combat this effect, researchers have targeted both encoding and retrieval processes. Enhanced encoding can reduce the misinformation effect (e.g., Lane, 2006; Pezdek & Roe, 1995), as can increasing memory monitoring at test by requiring participants to specify the source of reported details via a source-monitoring test (e.g., Lindsay & Johnson, 1989).

Practice at retrieving an event may provide a practical method for protecting memory from the influence of misinformation, given that encoding factors likely cannot be controlled in eyewitness situations. There are good reasons to expect that initial testing might benefit memory accuracy. For one, testing has been shown to generate 'mediator' memory traces that can later serve as effective retrieval cues (Carpenter, 2011; Pyc & Rawson, 2010) and can also en-

hance memory organization (Congleton & Rajaram, 2012). Testing can also selectively increase the memory strength of retrieved items, reducing their rate of forgetting (Kornell, Bjork, & Garcia, 2011), and can also facilitate accurate retrieval by enhancing memory for source information (Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010; Chan & McDermott, 2007). Thus, initial testing may improve the initial encoding of an event and later memory monitoring at test.

Consistent with these beneficial effects of testing, some studies have found that initial testing reduces the misinformation effect. Loftus (1977) showed participants a slideshow depicting a green car driving past an auto accident. Asking participants to indicate the car's color prior to exposure to a misleading suggestion that the car was blue reduced the misinformation effect on a final test (see also Loftus, 1979). Recently, Pansky and Tenenboim (2011) reported that initial testing of highly specific verbatim details of a witnessed crime reduced suggestibility relative to initial testing of broad gist-based details (cf. Lane, Mather, Villa, & Morita, 2001). Likewise, Memon, Zaragoza, Clifford, and Kidd (2010) found that completing the cognitive interview before (versus after) exposure to misleading details also reduced suggestibility (see also Gabbert, Hope, Fisher, & Jamieson, 2012).

Consistent with these beneficial effects of initial testing, Huff, Davis, and Meade (2013) reported a reduction in misinformation effects using a *social-contagion-of-memory paradigm* in which misinformation is introduced via an implied social source (e.g., McNabb & Meade, 2014; Meade & Roediger, 2002; Roediger, Meade, & Bergman, 2001), as opposed to another participant or confederate (e.g., Bodner, Musch, & Azad, 2009; Gabbert, Memon, & Allan, 2003; Hoffman, Granhag, See, & Loftus, 2001). Participants viewed a series of household scenes (e.g., bathroom and bedroom) each containing a variety of typical objects. They then reviewed a set of fake recall tests ostensibly completed by previous participants. Embedded within these tests were non-presented 'contagion items' that were schematically consistent with a given scene. Participants were exposed to

*Correspondence to: Mark J. Huff, Department of Psychology, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri, 63130, USA.
E-mail: mhuff@wustl.edu, mjhuff16@gmail.com

each contagion item zero, one, or four times across the set of fake recall tests. Critically, half of the participants completed an initial recall test after viewing the scenes but before the misinformation was introduced. Initial testing did not affect reporting of the contagion items on a final free recall test. However, on a final source-monitoring test, initial testing made participants less likely to falsely attribute contagion items to the scenes—a *protective effect of testing* (PET). Although the misinformation effect was stronger following four exposures than one exposure to contagion items, initial testing was equally effective at reducing suggestibility across exposures. In other words, the effectiveness of initial testing was not contingent on the strength of the misinformation. Initial testing therefore appears to improve memory accuracy, at least when misinformation is supplied by a social source—which is a very common potential source of influence in actual eyewitness situations (Paterson & Kemp, 2006).

Surprising then, is a set of demonstrations beginning with Chan, Thomas, and Bulevich (2009), in which initial testing *increased* suggestibility—a phenomenon dubbed *retrieval-enhanced suggestibility* (RES). In the paradigm of Chan et al., after a to-be-remembered event (e.g., an episode of the television show *24*), an initial test group completed a cued recall test about specific details in the episode, whereas a no-test group did not. Both groups were then exposed to misleading details about the event through an experimenter-prepared narrative summary. On a final cued recall test, misleading details were more likely to be reported by the initial test group than a no-test group (see also Chan & Langley, 2011; Chan & LaPaglia, 2011; Thomas, Bulevich, & Chan, 2010). This RES pattern has been shown whether testing is completed immediately or after a delay (Chan & LaPaglia, 2011) whether the initial test is cued or free recall (Wilford, Chan, & Tuhn, 2014) and persists when the final test requires participants to specify contextual details via a source-monitoring test (Chan, Wilford, & Hughes, 2012).

However, LaPaglia and Chan (2013) demonstrated that initial testing can produce a PET pattern in this paradigm if misinformation is presented via misleading questions rather than a narrative. Whether initial testing yields a RES pattern or PET pattern may thus be contingent, in part, on how the initial test shapes the learning of subsequent misinformation (Gordon, Thomas, & Bulevich, 2015). For example, Gordon and Thomas (2014) found evidence that initial testing directs attention to misleading details within the post-event information (see also Tousignant, Hall, & Loftus, 1986). Specifically, the RES pattern was associated with longer reading times for misinformation in a narrative, suggesting the misleading details received additional processing that enhanced learning and subsequent reporting of these items on a final test. In the social contagion paradigm of Huff et al. (2013), in contrast, the misinformation (contagion items) were always additive (i.e., not in the scenes) rather than contradictory (i.e., contradicting specific objects in the scenes; Frost, 2000; Nemeth & Belli, 2006). Therefore, discrepancies between the original information and misinformation may need to be present to trigger the additional processing of misinformation that yields the RES pattern.

Our study aimed to establish whether initial testing reliably improves memory accuracy (i.e., a PET pattern) in the social contagion paradigm, given the preponderance of the RES pattern in nonsocial misinformation paradigms. In particular, we sought to determine whether a PET pattern could be obtained in free recall (cf. Huff et al., 2013). This question is important from an applied perspective, given that free recall shares similar characteristics with the cognitive interview used in forensic settings (Fisher & Geiselman, 1992). Asking eyewitnesses to begin their accounts with free recall may thus benefit memory accuracy (Wilford et al., 2014), even though this procedure is not universally used in practice (Brunel & Py, 2013; Wells, Memon, & Penrod, 2006). Finding a PET in free recall would broaden the evidence that initial testing sometimes improves memory accuracy and thus would provide further incentive for studying the application of initial free recall techniques in forensic settings.

To help delineate the conditions that yield a PET pattern, we used two manipulations that have increased the beneficial effects of testing on correct memory in other paradigms. First, we evaluated whether the PET pattern is enhanced when participants complete more than one initial recall test. Completing multiple memory tests has been found to improve memory accuracy relative to a single test (Karpicke & Roediger, 2007). Thus, if initial testing protects memory from misinformation by increasing correct memory, then increasing the number of initial tests should reduce misinformation effects by further increasing correct memory. To evaluate this possibility, our participants either completed zero, one, or two initial free recall tests.

Testing benefits have also been found to increase with delay, once more forgetting of the initial event has occurred (Roediger & Karpicke, 2006). If testing impedes forgetting, then initial testing should improve memory, which in turn may increase memory's resistance to misinformation. If so, the benefits of initial testing may increase in situations where misinformation is not encountered immediately following an event. In eyewitness situations, there is typically a gap between the event and reports (and between the event and subsequent testimony, of course). To evaluate this factor, our participants either completed their final memory tests in an immediate condition or a 2-day-delay condition.

In sum, participants viewed six slides depicting household scenes and then completed zero, one, or two initial free recall tests. Either immediately or following a 48-hour delay, they then completed the social contagion phase in which they reviewed a set of recall tests ostensibly completed by other participants to expose them to non-presented contagion items. The number of exposures to contagion items was also varied (zero, one, or four times) to determine whether initial testing effects are modulated by the magnitude of the misinformation effect. Participants then completed final free recall and source-memory tests. We expected that increasing the number of exposures to contagion items would increase false recall and false source attributions (Mitchell & Zaragoza, 1996). Based on the testing effect literature, we also expected that increasing the number of initial tests and/or the delay before misinformation and final testing in the procedure of Huff et al. (2013) would enhance the PET pattern on recall memory and source memory.

METHOD

Participants

University of Calgary undergraduates ($N = 216$; 36 per cell) participated for course credit. The immediate and delay conditions were tested in consecutive years across both fall and winter semesters, using participants recruited from the same research participation pool. Although participants were not randomly assigned to delay condition, delay was nonetheless treated as a random factor given the likely similarity in participant characteristics, and given that the same experimenter collected the data. Within each delay condition, participants were randomly assigned to the zero, one, or two initial test conditions. One participant was replaced for not following test instructions, and eight were replaced in the delay condition because of attrition. Participants spoke fluent English and had normal or corrected-to-normal vision.

Materials

We constructed digital color images depicting six household scenes (toolbox, bathroom, kitchen, bedroom, closet, and desk; after Huff et al., 2013; Meade & Roediger, 2002; see Figure 1 for an example). Each scene displayed objects ($M = 23.83$) frequently listed by 18 additional undergraduates who listed items they would expect to see in each scene. The two most frequently listed items for each scene (i.e., high-expectancy items) served as the contagion items (nails/screwdriver, soap/toothbrush, knives/plates, lamp/pillow, jacket/shoes, and paper/pens) and thus were not presented in the scene images we constructed.

Following Huff et al. (2013), fake recall tests were created to introduce contagion items to participants. Five colleagues handwrote one recall test for each of the six scenes. These recall tests were then photocopied and organized into packets of 30 recall tests ostensibly completed by five other participants from a previous experiment. Each test included 6–10 designated items. Contagion items were always written in serial positions four and six, and correct items were randomly placed in the remaining list positions. Recall sheets from one writer contained only correct items from the scene.
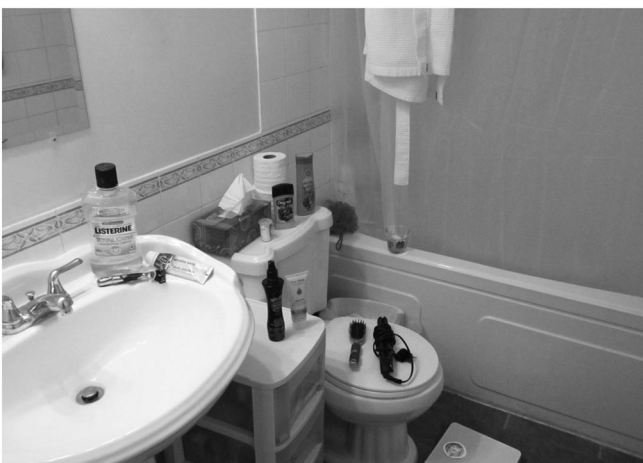

Figure 1. Sample household scene (soap and toothbrush were the non-presented contagion items)

Contagion items were provided from the four remaining writers. Exposures to contagion items were counterbalanced across the scenes such that of the six scenes, zero writers presented contagion items for each of two scenes (zero-exposure items), one writer presented contagion items for each of two scenes (one-exposure items), and four writers presented contagion items for each of two scenes (four-exposure items).

Procedure

Figure 2 depicts the design. Groups of up to six participants were tested. They were told they would view a series of household scenes, and their memory for the items in the scenes would later be tested. Intentional instructions were used under the assumption that eyewitnesses engage in intentional encoding in eyewitness situations. Each scene was presented on a large projector screen for 15 seconds in the order listed earlier and verbally labeled by an experimenter. Following study of the scenes, participants completed an arithmetic filler task for 2 minutes. The zero-test group performed this filler task for 12 additional minutes, whereas the one-test and two-test groups completed a free recall test for each scene. The scenes were tested in the order with which they were studied. Each of six sheets listed the scene name at the top, and participants had 2 minutes to recall its objects. Participants were instructed to 'write down as many items as you can remember from the scene listed at the top of page'. The two-test group then immediately recalled all six scenes a second time in the same order (2 minutes each). Thus, the initial test phase was 12 minutes for the one-test groups and 24 minutes for the two-test groups.

Participants in the immediate test condition continued with the contagion phase, whereas those in the delayed condition were dismissed and returned after 48 hours to begin the contagion phase (Figure 2). The contagion phase was modeled after Huff et al. (2013). Each participant received a packet containing five sets of six recall tests ostensibly completed by previous participants for another experiment. Participants were (falsely) told that a focus of the study was to determine how pleasantness influences memory for objects in the scenes. Participants were asked to review each recall test (presented in the order of the studied scenes) and to circle the objects they found pleasant. This pleasantness task was used to promote attention to the items in the fake tests.

Immediately after the contagion phase, participants completed a final 12-minute free recall test identical to the initial test procedure (six scenes, 2 minutes per scene). Recall was immediately followed by a 36-item source-monitoring recognition test: A sheet containing a random ordering of 18 correct items (three per scene), 12 contagion items (two per scene), and 6 novel filler items (not presented in the scenes or on the fake tests). Participants classified their memory for each item as *scene* (item was in the original scene), *other* (item was on the other participants' recall tests), *both* (item was in the original scene and on the other participants' recall tests), or *neither*. Finally, participants were probed for suspicion and for prior knowledge of the misinformation effect; none warranted replacement for these reasons.
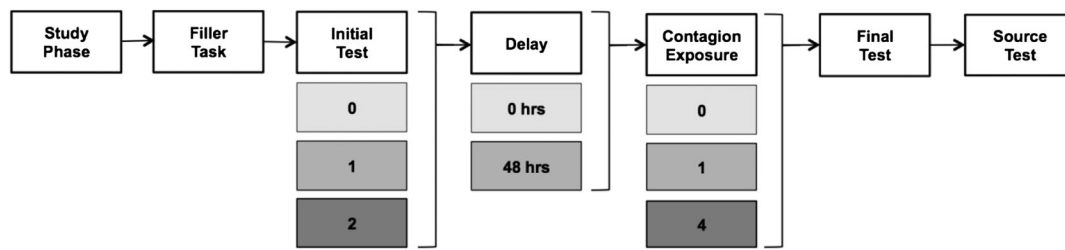
Figure 2. Study design

RESULTS

A $p < .05$ significance level was used except as noted. Effect sizes for significant comparisons were calculated using partial eta squared (η_p^2) for analyses of variance (ANOVAs), and Cohen's d for t-tests.

Free recall

Table 1 provides the proportion of objects from the scenes that were correctly recalled on each test. Correct recall was computed by dividing the number of items recalled in a given scene by the total number of items presented in a given scene. A lenient scoring criterion was adopted such that misspellings and synonyms of scene items (e.g., 'pan' would be counted for 'pot' for the kitchen scene) were both counted. The proportion of scene items recalled was then analyzed using a 3 (initial test: 0 vs. 1 vs. 2) × 2 (delay: 0 vs. 48 hours) between-subjects ANOVA. An effect of initial test was found, $F(2, 210) = 53.16$, $MSE = .01$, $\eta_p^2 = 0.09$. Confirming a retrieval-practice effect, correct recall was greater after both one and two initial tests relative to zero initial tests (0.36 vs. 0.30; 0.35 vs. 0.30), $t(142) = 3.69$, $SEM = .01$, $d = 0.62$, and $t(142) = 3.40$, $SEM = 0.01$, $d = 0.57$, respectively. However, taking two initial tests was not more beneficial than taking one (0.35 vs. 0.36), $t < 1$. Correct recall was lower after delay (0.30 vs. 0.38), $F(1, 210) = 53.16$, $MSE = 0.01$, $\eta_p^2 = 0.20$. The interaction was not significant, $F < 1$.

False recall of contagion items (Table 2) was calculated as the number of contagion items reported in a given scene divided by two and was scored as for correct recall. To help the reader gauge the magnitude of the contagion effects, and in keeping with past studies (Huff et al., 2013; Meade & Roediger, 2002), Table 2 also provides corrected contagion scores computed by subtracting the zero-exposure condition from the one- and four-exposure conditions.

The proportion of contagion items recalled was analyzed in a 3 (exposure: 0 vs. 1 vs. 4) × 3 (initial test: 0 vs. 1 vs. 2) × 2 (delay: immediate vs. 48 hours) mixed-factor ANOVA. The effect of contagion exposure, $F(2, 420)$

$= 109.09$, $MSE = 0.05$, $\eta_p^2 = 0.34$, reflected greater contagion recall after one than zero exposures (0.33 vs. 0.18), $t(215) = 7.47$, $SEM = 0.02$, $d = 0.65$, after four than one exposures (0.50 vs. 0.33), $t(215) = 7.13$, $SEM = 0.02$, $d = 0.63$, and after four than zero exposures (0.50 vs. 0.18), $t(215) = 15.09$, $SEM = 0.02$, $d = 1.32$. An effect of initial testing was also found, $F(2, 210) = 9.61$, $MSE = 0.08$, $\eta_p^2 = 0.08$. Contagion item recall was reduced after one than zero initial tests (0.30 vs. 0.41), $t(142) = 3.90$, $SEM = 0.02$, $d = 0.65$, and after two than zero initial tests (0.31 vs. 0.41), $t(142) = 3.53$, $SEM = 0.02$, $d = 0.59$. However, contagion item recall was similar after one or two initial tests (0.30 vs. 0.31), $t < 1$. The effect of delay was not reliable, $F(1, 210) = 2.79$, $MSE = 0.08$, $p = .10$, nor was the interaction of exposure and initial test, $F(4, 420) = 2.22$, $MSE = 0.05$, $p = .07$.

Critically, the effect of initial testing on contagion recall interacted with delay, $F(2, 210) = 6.44$, $MSE = 0.08$, $\eta_p^2 = 0.06$. In the immediate condition, contagion recall was not significantly lower after one than zero tests (0.31 vs. 0.38), $t(70) = 1.68$, $SEM = 0.03$, $p = .10$, after two versus zero tests (0.38 vs. 0.38), $t < 1$, or after two or one tests (0.38 vs. 0.31), $t(70) = 1.56$, $SEM = 0.03$, $p = .12$. In contrast, in the delayed condition, contagion recall was lower after one than zero tests (0.28 vs. 0.43), $t(70) = 3.85$, $SEM = 0.03$, $d = 0.92$, lower after two than zero tests (0.24 vs. 0.43), $t(70) = 5.54$, $SEM = 0.03$, $d = 1.32$, but was again equivalent after two or one tests (0.24 vs. 0.28), $t(70) = 1.31$, $SEM = 0.03$, $p = .19$. Thus, taking either one or two initial tests yielded a robust PET pattern on recall—the first time a PET has been reported on this memory test. The remaining interactions were not significant, $Fs < 1$.

Source monitoring

Source misattributions for contagion items were operationalized as the proportion of contagion items that were misattributed to the scenes (see Table 3, 'Contagion effect' row). These misattributions were analyzed as in contagion recall. The main effect of exposure, $F(2, 420) = 25.68$, $MSE = 0.06$, $\eta_p^2 = 0.11$, reflected an increase in misattributions after one than zero exposures (0.55 vs. 0.47), $t(215)$

Table 1. Mean (SD) proportion of correct recall of scene items on each test

Test/group	Immediate test			Delayed test		
	0 Initial test	1 Initial test	2 Initial test	0 Initial test	1 Initial test	2 Initial test
Initial test 1	—	0.30 (0.08)	0.28 (0.08)	—	0.28 (0.07)	0.26 (0.06)
Initial test 2	—	—	0.29 (0.10)	—	—	0.27 (0.06)
Final test	0.34 (0.08)	0.39 (0.08)	0.39 (0.09)	0.26 (0.07)	0.32 (0.07)	0.31 (0.06)

Table 2. Mean (SD) proportion of contagion items recalled, and corrected contagion effects on the final recall test

	Immediate test			Delayed test		
	0 Initial test	1 Initial test	2 Initial test	0 Initial test	1 Initial test	2 Initial test
Exposure condition						
0 Exposure	0.17 (0.20)	0.17 (0.19)	0.26 (0.24)	0.24 (0.20)	0.13 (0.15)	0.13 (0.17)
1 Exposure	0.42 (0.31)	0.32 (0.27)	0.35 (0.23)	0.40 (0.25)	0.27 (0.19)	0.23 (0.27)
4 Exposure	0.54 (0.26)	0.45 (0.27)	0.51 (0.27)	0.65 (0.27)	0.46 (0.31)	0.35 (0.24)
Corrected contagion effect						
1 Exposure	0.24 (0.33)	0.15 (0.32)	0.08 (0.31)	0.16 (0.29)	0.15 (0.17)	0.10 (0.28)
4 Exposure	0.37 (0.33)	0.28 (0.32)	0.25 (0.27)	0.41 (0.32)	0.33 (0.28)	0.23 (0.28)

Note: In the delayed test condition, misinformation was presented after a 48-hour delay, followed by the final recall and source tests. Corrected contagion effects were computed by subtracting the zero exposures proportion from the one and four exposures proportion for each participant; they were provided to help the reader assess the effects of exposure relative to baseline and were not analyzed.

Table 3. Mean (SD) proportion of source attributions for contagion items

Exp./source attribution	0 Initial test			1 Initial test			2 Initial test		
	0 Exp.	1 Exp.	4 Exp.	0 Exp.	1 Exp.	4 Exp.	0 Exp.	1 Exp.	4 Exp.
Immediate test									
'Scene'	0.27 (0.29)	0.19 (0.22)	0.13 (0.19)	0.20 (0.23)	0.15 (0.23)	0.10 (0.25)	0.24 (0.28)	0.20 (0.23)	0.10 (0.19)
'Scene and other'	0.26 (0.25)	0.44 (0.30)	0.64 (0.25)	0.17 (0.23)	0.31 (0.31)	0.46 (0.34)	0.30 (0.29)	0.36 (0.28)	0.52 (0.34)
Contagion effect	**0.53 (0.32)**	**0.63 (0.30)**	**0.77 (0.24)**	**0.37 (0.28)**	**0.46 (0.32)**	**0.56 (0.32)**	**0.53 (0.26)**	**0.56 (0.29)**	**0.62 (0.34)**
'Other only'	0.11 (0.18)	0.22 (0.29)	0.22 (0.24)	0.18 (0.24)	0.30 (0.21)	0.34 (0.30)	0.16 (0.18)	0.31 (0.30)	0.31 (0.28)
'Neither'	0.36 (0.31)	0.15 (0.23)	0.02 (0.07)	0.45 (0.29)	0.24 (0.28)	0.10 (0.15)	0.30 (0.29)	0.14 (0.16)	0.07 (0.15)
Delayed test									
'Scene'	0.34 (0.33)	0.25 (0.29)	0.08 (0.22)	0.20 (0.23)	0.16 (0.23)	0.10 (0.23)	0.16 (0.21)	0.17 (0.25)	0.03 (0.09)
'Scene and other'	0.27 (0.23)	0.52 (0.28)	0.73 (0.30)	0.22 (0.22)	0.28 (0.26)	0.44 (0.32)	0.20 (0.22)	0.26 (0.27)	0.47 (0.29)
Contagion effect	**0.59 (0.30)**	**0.77 (0.25)**	**0.81 (0.23)**	**0.42 (0.23)**	**0.44 (0.30)**	**0.54 (0.32)**	**0.36 (0.26)**	**0.42 (0.34)**	**0.50 (0.31)**
'Other only'	0.15 (0.21)	0.13 (0.18)	0.18 (0.23)	0.14 (0.16)	0.32 (0.27)	0.38 (0.30)	0.24 (0.25)	0.34 (0.30)	0.43 (0.31)
'Neither'	0.26 (0.23)	0.10 (0.15)	0.01 (0.05)	0.44 (0.27)	0.25 (0.20)	0.08 (0.15)	0.40 (0.25)	0.24 (0.26)	0.07 (0.13)

Note: In the delayed test condition, misinformation was presented after a 48-hour delay, followed by the final recall and source tests. The contagion effect row (in bold) is the sum of the 'Scene' and 'Scene and other' rows; it captures the total proportion of contagion items that were misattributed to the scenes.

= 3.47, $SEM = 0.02$, $d = 0.26$, after four than one exposures (0.63 vs. 0.55), $t(215) = 3.76$, $SEM = 0.02$, $d = 0.25$, and after four than zero exposures (0.63 vs. 0.47), $t(215) = 7.07$, $SEM = 0.02$, $d = 0.53$. Importantly, a main effect of initial test was also found, $F(2, 210) = 21.82$, $MSE = 0.14$, $\eta_p^2 = 0.17$. Misattributions were less frequent after one than zero tests (0.46 vs. 0.68), $t(142) = 6.27$, $SEM = 0.03$, $d = 1.05$, and after two than zero tests (0.50 vs. 0.68), $t(142) = 5.10$, $SEM = 0.03$, $d = 0.86$. However, as was true of contagion item recall, misattributions were similar after one or two initial tests (0.46 vs. 0.50), $t < 1$. The main effect of delay was not reliable, $F < 1$.

As was true of recall of contagion items, the effect of initial testing on source judgments interacted with delay, $F(2, 210) = 21.82$, $MSE = 0.14$, $\eta_p^2 = 0.05$. Figure 3 captures this interaction. In the immediate test condition, misattributions were lower after one than zero tests (0.46 vs. 0.64), $t(70) = 3.55$, $SEM = 0.04$, $d = 0.85$, but only numerically so after two than zero tests (0.57 vs. 0.64), $t(70) = 1.44$, $SEM = 0.04$, $p = .16$. Unexpectedly, misattributions were marginally more common after two than one initial test (0.57 vs. 0.46), $t(70) = 2.03$, $SEM = 0.04$, $p = .05$, $d = 0.49$. In the delay test condition, taking either one or two tests reduced misattributions relative to the zero-test group (0.43 vs.

0.73, 0.46 vs. 0.73), $t(70) = 5.36$, $SEM = 0.04$, $d = 1.28$, and $t(70) = 6.07$, $SEM = 0.04$, $d = 1.45$, respectively, whereas misattributions were equivalent after one or two tests (0.43 vs. 0.46), $t < 1$. The remaining interactions did not reach significance, $Fs > 1.41$, $ps > .20$.

Correct attributions for contagion items (see Table 3, 'Other' row) were subject to the same analysis. There was
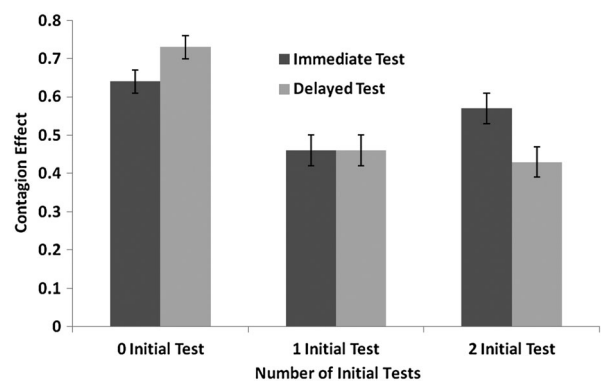

Figure 3. Proportion of contagion effect source misattributions ('Scene' and 'Scene and other' attributions) for contagion items for initial test and immediate and delayed test groups collapsed across exposures. Bars reflect standard error

again an effect of contagion exposure, $F(2, 420) = 23.40$, $MSE = 0.05$, $\eta_p^2 = 0.10$. Contagion items were more likely to be correctly attributed to 'other participants' after one than zero exposures (0.16 vs. 0.27), $t(215) = 5.05$, $SEM = 0.02$, $d = 0.46$, and after four than zero exposures (0.31 vs. 0.16), $t(215) = 6.30$, $SEM = 0.02$, $d = 0.60$, although the difference for four versus one exposures was only marginal (0.31 vs. 0.27), $t(215) = 1.78$, $SEM = 0.02$, $p = .08$, $d = 0.14$. The effect of initial test, $F(2, 210) = 12.73$, $MSE = 0.09$, $\eta_p^2 = 0.11$, reflected more correct attributions after one than zero tests (0.28 vs. 0.17), $t(142) = 4.11$, $SEM = 0.02$, $d = 0.69$, and after two than zero tests (0.30 vs. 0.17), $t(142) = 4.81$, $SEM = 0.02$, $d = 0.81$, but not after two than one tests (0.30 vs. 0.28), $t < 1$. The effect of delay was not significant, $F < 1$. The interactions, including the interaction between initial test and delay, did not reach significance ($Fs < 1.88$, $ps > .15$).

Finally, correct attributions for scene items (see Table 4, Total correct rows) were also analyzed as earlier. Here, unexpectedly, the effect of initial test, $F(2, 210) = 13.84$, $MSE = 0.03$, $\eta_p^2 = 0.12$, reflected *fewer* correct attributions after one than zero tests (0.51 vs. 0.63), $t(142) = 3.87$, $SEM = 0.02$, $d = 0.65$, after two than zero tests (0.49 vs. 0.63), $t(142) = 4.87$, $SEM = 0.02$, $d = 0.82$, but equivalent rates after one or two tests (0.51 vs. 0.49), $t(142) = 1.15$, $SEM = 0.02$, $p = .25$. The effect of delay and the interaction did not reach significance ($Fs < 2.27$, $ps > .13$).

DISCUSSION

Using the social-contagion-of-memory paradigm developed by Roediger et al. (2001), we explored how initial memory testing affects later suggestibility to misinformation. Misleading information was provided by an implied social source—the review of recall sheets ostensibly from other participants that included non-studied 'contagion' items. We replicated the finding of a PET pattern on source monitoring by Huff et al. (2013): Initial testing made participants less likely to falsely attribute contagion items to the study

Table 4. Mean (*SD*) proportion of source attributions for correct items

Tests/Source attributions	0 Initial test	1 Initial test	2 Initial test
Immediate test			
'Scene'	0.27 (0.15)	0.17 (0.17)	0.16 (0.17)
'Scene and other'	0.42 (0.17)	0.35 (0.20)	0.35 (0.19)
Total correct	**0.66 (0.16)**	**0.52 (0.17)**	**0.51 (0.18)**
'Other only'	0.19 (0.14)	0.27 (0.18)	0.26 (0.14)
'Neither'	0.15 (0.09)	0.21 (0.15)	0.23 (0.14)
Delayed test			
'Scene'	0.21 (0.19)	0.17 (0.15)	0.15 (0.10)
'Scene and other'	0.40 (0.18)	0.34 (0.17)	0.31 (0.14)
Total correct	**0.61 (0.21)**	**0.51 (0.17)**	**0.46 (0.15)**
'Other only'	0.22 (0.17)	0.27 (0.16)	0.32 (0.16)
'Neither'	0.17 (0.14)	0.21 (0.12)	0.22 (0.12)

Note: In the delayed test condition, misinformation was presented after a 48-hour delay, followed by the final recall and source tests. The total correct row (in bold) is the sum of the 'Scene' and 'Scene and other' rows, which captures the total proportion of studied objects that were correctly attributed to the scene.

scenes. This pattern was found regardless of whether exposure to contagion items occurred immediately after initial testing or was delayed 48 hours. An important and novel finding was that delayed exposure to contagion items also produced a PET pattern on free recall: Initial testing made participants less likely to freely report contagion items. We suggest that initial testing benefitted recall after a delay by slowing forgetting of the scenes (Roediger & Karpicke, 2006). When forgetting is greater, as is the case after a delay, initial testing can reduce suggestibility effects in free recall. Moreover, the PET pattern on delayed recall, and on source monitoring at both retention intervals, was similar whether contagion items were suggested one or four times. Initial testing thus appears to reduce false memory similarly for misinformation of varying strength. In contrast, taking two initial tests did not increase the PET pattern on either memory test beyond the benefits obtained from taking one initial test. Below we consider the theoretical and applied implications of our findings. We also consider why initial testing yielded beneficial effects on memory in our paradigm, whereas it often *increases* misinformation effects (i.e., the RES pattern) in other paradigms.

We first consider why taking two (versus one) initial recall tests failed to yield a larger PET pattern. In the testing effect literature, repeated initial testing is more effective when tests are spaced over equal intervals rather than massed (Karpicke & Roediger, 2007). This spacing advantage occurs whether the final test is completed after a short (10 minutes) or long (48 hours) retention interval, similar to the intervals we used. In our experiment, the two-test group completed their pair of initial recall tests consecutively, which may not have introduced sufficient spacing to strengthen the effects of initial testing. The possibility that completing more than one initial test increases the PET pattern if the initial tests are spaced remains to be tested.

Although initial testing generally benefitted memory accuracy, we also found some potential costs of initial testing. First, in the immediate test condition, contagion items were marginally more likely to be attributed to the scenes after two (versus one) initial tests. We hesitate to place much weight on this finding given it was unreliable and did not replicate in the delayed condition. Second, initial testing reduced how often scene items were correctly attributed to the scenes, a finding also reported by Huff et al. (2013, Experiment 3). One possibility is that the initial recall tests may have led participants to deem their memories for the scenes to be poor, thus leading them to adopt a more conservative response criterion for attributing items to the scenes on the source-monitoring test. Consistent with this possibility, 'neither' attributions for scene items were greater in both the one-test (0.21) and two-test groups (0.22) than the zero-test group (0.16), $t(142) = 2.37$, $SEM = 0.01$, and $t(142) = 3.02$, $SEM = 0.01$ (Table 4). If poor performance on the initial tests induced a conservative response bias, we may have underestimated the potential benefits of initial testing on the final source test. However, this cost was not found in free recall, where initial testing instead benefitted recall on both immediate and delayed tests. Regardless, these 'costs' of initial testing are a reminder of the importance of examining how manipulations influence both false and correct memory in

false memory paradigms (e.g., Gunter, Ivanko, & Bodner, 2005; Huff et al., 2015).

The PET patterns we obtained are the opposite of the frequently reported RES pattern (Chan et al., 2009; 2012; Chan & Langley, 2011; Chan & LaPaglia, 2011; LaPaglia & Chan, 2013; Thomas et al., 2010; Wilford et al., 2014). An important area for future research will be to determine when initial testing is likely to have protective (PET) versus harmful (RES) effects on memory. One determining factor appears to be whether the initial test directs attention towards misinformation, which can increase encoding of misinformation (Gordon & Thomas, 2014; Gordon et al., 2015). To date, the PET pattern has only been tested with additive misinformation, whereas the RES pattern has only been tested following contradictory misinformation. Therefore, initial testing may typically increase suggestibility for contradictory details but decrease suggestibility for additive details. One exception to this pattern is LaPaglia and Chan (2013), who found a PET pattern when contradictory misinformation was embedded in misleading questions rather than in a narrative (which yielded the usual RES pattern). Participants do not answer misleading questions with the misleading details, and therefore, attention is not as likely to be directed to the misleading details as in the case of a narrative. Additive misinformation, as presented in our social contagion phase, may have operated similarly to misleading questions in this respect given the absence of a detectable contradiction.

Importantly, whether the RES pattern occurs when misinformation is provided by a social source has not been investigated. Thus, it remains possible that initial testing might generally be beneficial when misinformation was introduced by a social source, as is common in eyewitness situations. In traditional misinformation experiments (and in all studies reporting the RES pattern), misinformation is presented via experimenter-prepared materials such as detailed summaries to which eyewitnesses are unlikely to be exposed. A participant may be more likely to adopt misinformation when presented from experimenter-prepared sources because of an expectation that the experimental materials are accurate (McCloskey & Zaragoza, 1985). In contrast, participants likely deem memory information provided by social sources as fallible, as is true of their own memory, and therefore may deem that information less credible. Consistent with this possibility, misinformation effects are often stronger when the misinformation is presented by a more credible source (e.g., Underwood & Pezdek, 1998). Completing an initial test may increase participants' reporting of misinformation from a source they deem to be reliable and trustworthy. Thus, misinformation presented through a social source may be a better approximation of suggestibility in actual eyewitness situations.

Other factors likely also contribute to whether a PET pattern or RES pattern occurs. Such factors include the type of study event (images of household scenes versus television episodes) and the method used to introduce misinformation (fake recall tests versus misleading questions versus a narrative summary). Determining the influence of such factors should help inform guidelines for the use of initial free recall testing when interviewing eyewitnesses.

Finally, our paradigm was not intended to mimic actual 'eyewitness' situations in terms of materials (e.g., household scenes versus crime scenes, fake recall tests versus misinformation from other witnesses); however, it shares many elements found in eyewitness scenarios. For example, misinformation encountered in eyewitness events is more likely to affect peripheral objects and details rather than central 'narrative' aspects (Heath & Erickson, 1998; Wilford et al., 2014) akin to the social contagion paradigm. Further, misinformation encountered socially is likely a common source in eyewitness events (Paterson & Kemp, 2006)—perhaps more so than exposure to detailed experimenter-prepared narratives generally used in misinformation paradigms. Finally, using the social contagion paradigm with fake recall tests also provides an important similarity to eyewitness events in terms of the number of exposures to misinformation. Eyewitnesses are likely exposed to misinformation multiple times prior to providing their testimony. By manipulating the number of exposures to suggested details, the social contagion paradigm provides a simple means of contrasting the effects of single versus repeated exposures on suggestibility. Thus, our social contagion paradigm shares many similarities with eyewitness events and other misinformation paradigms and even offers advantages that may be useful for studying factors influencing eyewitness memory.

CONCLUSION

How initial memory testing modulates the effects of exposure to misleading information contributes to our understanding of memory. These findings also have the potential to contribute to guidelines for interviewing eyewitnesses, as well as for the interpretation of testimony in legal contexts. In our work using the social-contagion-of-memory paradigm (present study; Huff et al., 2013), initial testing has typically had protective effects on memory, rather than increasing the misinformation effect. However, delineating the conditions under which initial testing decreases suggestibility remains an important direction for future research, given other evidence that initial testing can increase the misinformation effect (e.g., Chan et al., 2009).

REFERENCES

Bodner, G. E., Musch, E., & Azad, T. (2009). Reevaluating the potency of the memory conformity effect. *Memory & Cognition*, *37*, 1069–1076. DOI:10.3758/MC.37.8.1069.

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.

Brunel, M., & Py, J. (2013). Questioning the acceptability of the cognitive interview to improve its use. *L'Année Psychologique*, *113*, 427–458. DOI:10.4074/S0003503313003059.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, 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.

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., & 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). Recall 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*, *67*, 78–85. DOI:10.1016/j.jml.2012.02.006.

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

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, England: Charles C. Thomas.

Frost, P. (2000). The quality of false memory over time: Is memory for misinformation 'remembered' or 'known'? *Psychonomic Bulletin & Review*, *7*, 531–536. DOI:10.3758/BF03214367.

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.

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology*, *17*, 533–543. DOI:10.1002/acp.885.

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, *8*, 579–586. DOI:10.3758/BF03196194.

Gordon, L. T., & Thomas, A. K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory & Cognition*, *42*, 186–197. DOI:10.3458/s13421-013-0361-2.

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 & Language*, *83*, 140–151. DOI:10.1016/j.jml.2015.03.007.

Gunter, R. W., Ivanko, S. L., & Bodner, G. E. (2005). Can test list context manipulations improve recognition accuracy in the DRM paradigm? *Memory*, *13*, 862–873. DOI:10.1080/09658210444000458.

Heath, W. P., & Erickson, J. R. (1998). Memory for central and peripheral actions and props after varied post-event presentation. *Legal and Criminological Psychology*, *3*, 321–346. DOI:10.1111/j.2044-8333-1998.tb00369.x.

Hoffman, H. G., Granhag, P. A., See, S. T. K., & Loftus, E. F. (2001). Social influences on reality-monitoring decisions. *Memory & Cognition*, *29*, 394–404. DOI:10.3758/BF03196390.

Huff, M. J., Bodner, G. E., & Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: A meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psychonomic Bulletin & Review*, *22*, 321–346. DOI:10.3758/s13423-014-0648-8.

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.

Hunt, R. R., & Worthen, J. B. (2006). *Distinctiveness and memory*. New York: Oxford University Press.

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 704–719. DOI:10.1016/j.jml.2006.09.004.

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., & 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.

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 a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 19–31. DOI:10.1037/0878-7393.4.1.19.

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

McNabb, J. C., & Meade, M. L. (2014). Correcting socially introduced false memories: The effect of restudy. *Journal of Applied Research in Memory and Cognition*, *3*, 287–292. DOI:10.1016/j.jarmac.2015.05.007.

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., & Zaragoza, M. S. (1996). Repeated exposure to suggestion and false memory: The role of contextual variability. *Journal of Memory and Language*, *35*, 246–260. DOI:10.1006/jmla.1996.0014.

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 Cognitive Psychology*, *20*, 563–573. DOI:10.1002/acp.1207.

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.

Paterson, H. M., & Kemp, R. I. (2006). Comparing methods of encountering post-event information: The power of co-witness suggestion. *Applied Cognitive Psychology*, *20*, 1083–1099. DOI:10.1002/acp.1261.

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

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/a0023956.

Roediger, H. L. III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. DOI:10.1111/j.1745-6916.2006.00012.x.

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

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.

Underwood, J., & Pezdek, K. (1998). Memory suggestibility as an example of the sleeper effect. *Psychonomic Bulletin & Review*, *5*, 449–453. DOI:10.1037/BF03208820.

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. 5–64). Mahwah: Erlbaum.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45–75. DOI:10.1111/j.1529-1006.2006.00027.x.