12-2011

The Dark Side of Testing Memory: Repeated Retrieval Can Enhance Eyewitness Suggestibility

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

Jessica A. LaPaglia
Iowa State University

The Dark Side of Testing Memory: Repeated Retrieval Can Enhance Eyewitness Suggestibility

Abstract

Eyewitnesses typically recount their experiences many times before trial. Such repeated retrieval can enhance memory retention of the witnessed event. However, recent studies (e.g., Chan, Thomas, & Bulevich, 2009) have found that initial retrieval can exacerbate eyewitness suggestibility to later misleading information—a finding termed retrieval-enhanced suggestibility (RES). Here we examined the influence of multiple retrieval attempts on eyewitness suggestibility to subsequent misinformation. In four experiments, we systematically varied the number of initial tests taken (between zero and six), the delay between initial testing and misinformation exposure (~30 min or 1 week), and whether initial testing was manipulated between- or within-subjects. University undergraduate students were used as participants. Overall, we found that eyewitness suggestibility increased as the number of initial tests increased, but this RES effect was qualified by the delay and by whether initial testing occurred in a within- or between-subjects manner. Specifically, the within-subjects RES effect was smaller than the between-subjects RES effect, possibly because of the influence of retrieval-induced forgetting/facilitation (Chan, 2009) when initial testing was manipulated within subjects. Moreover, consistent with the testing effect literature (Roediger & Karpicke, 2006), the benefits of repeated testing on later memory were stronger after a 1-week delay than after a 30-min delay, thus reducing the negative impact of RES in long-term situations. These findings suggest that conditions that are likely to occur in criminal investigations can either increase (repeated testing) or reduce (delay) the influence of RES, thus further demonstrating the complex relationship between eyewitness memory and repeated retrieval.

Disciplines

Cognitive Psychology | Evidence

Comments

This is a manuscript of an article from Journal of Experimental Psychology: Applied 17 (2011): 418, doi:10.1037/a0025147. Posted with permission. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

The Dark Side of Testing Memory:

Repeated Retrieval Can Enhance Eyewitness Suggestibility

Jason C. K. Chan

Jessica A. LaPaglia

Iowa State University of Science & Technology

(Word Count: 9471)

Address correspondence to:

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

ckchan@iastate.edu
(515) 294-6946

This is a manuscript of an article from *Journal of Experimental Psychology: Applied* 17 (2011): 418, doi:10.1037/a0025147. Posted with permission. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Abstract

Eyewitnesses typically recount their experiences many times before trial. Such repeated retrieval can enhance memory retention of the witnessed event. However, recent studies (e.g., Chan, Thomas, & Bulevich, 2009) have found that initial retrieval can exacerbate eyewitness suggestibility to later misleading information – a finding termed retrieval-enhanced suggestibility (RES). Here we examined the influence of multiple retrieval attempts on eyewitness suggestibility to subsequent misinformation. In four experiments, we systematically varied the number of initial tests taken (between zero and six), the delay between initial testing and misinformation exposure (~30 min or 1 week), and whether initial testing was manipulated between- or within-subjects. University undergraduate students were used as participants. Overall, we found that eyewitness suggestibility increased as the number of initial tests increased, but this RES effect was qualified by the delay and by whether initial testing occurred in a within- or between-subjects manner. Specifically, the within-subjects RES effect was smaller than the between-subjects RES effect, possibly due to the influence of retrieval-induced forgetting/facilitation (Chan, 2009) when initial testing was manipulated within-subjects. Moreover, consistent with the testing effect literature (Roediger & Karpicke, 2006), the benefits of repeated testing on later memory were stronger following a 1-week delay than following a 30 min delay, thus reducing the negative impact of RES in long-term situations. These findings suggest that conditions that are likely to occur in criminal investigations can either increase (repeated testing) or reduce (delay) the influence of RES, thus further demonstrating the complex relationship between eyewitness memory and repeated retrieval.

(Word Count: 249)

The Dark Side of Testing Memory: Repeated Retrieval Can Enhance Eyewitness Suggestibility

The legal process is long and arduous. After witnessing a crime, it is not unusual for weeks or months to pass before one provides testimony in court. For example, a six-month delay separated the murders and the beginning of the criminal trial in the highly publicized O.J. Simpson case. During the delay between witnessing a crime and providing testimony, witnesses are often questioned by police investigators, lawyers, friends, and family members (Gray, 1993; Whitcomb, Shapiro, & Stellwagen, 1985). Moreover, during this delay, eyewitnesses can encounter misleading information from various sources (Allen & Lindsay, 1998; Lindsay, Allen, Chan, & Dahl, 2004). It is well known that misinformation can degrade the later accuracy of eyewitness reports (for a recent review, see Zaragoza, Belli, & Payment, 2007). In the present study, we sought to examine the effects of repeated retrieval on the ability for people to resist subsequently presented misleading information.

The Effects of Repeated Retrieval on Retention

Research on the suggestibility of human memory has had a long tradition, beginning with the pioneering though underappreciated work of Binet (1900). In recent years, a controversy has emerged in the literature on children's memory. While some researchers assert that recall accuracy can decline over repeated retrieval attempts, especially if the interviews are conducted in a suggestive manner, others suggest the opposite, especially when open-ended, nonleading questions (as opposed to yes/no questions) are asked (for reviews on the topic, see Goodman & Quas, 2008; Hershkowitz & Terner, 2007; La Rooy, Katz, Malloy, & Lamb, 2010; La Rooy, Lamb, & Pipe, 2009; Memon & Vartoukian, 1996).

The findings in the adult memory literature are similarly mixed. While some studies showed that repeated retrieval can increase errors in memory reports (e.g., Bartlett, 1932;

Henkel, 2004; Hyman & Pentland, 1996; Roediger, Jacoby, & McDermott, 1996; Roediger & McDermott, 1995), others have shown the reverse (e.g., Ballard, 1913; Bluck, Levine, & Laulhere, 1999; Bornstein, Liebel, & Scarberry, 1998; Brown, 1923; Erdelyi & Becker, 1974; Scrivner & Safer, 1988). How can one reconcile these disparate findings? Wheeler and Roediger (1992) proposed that one deciding factor is the delay between successive recall attempts. Specifically, if the delay between recall trials is short, accurate recall probability is likely to rise between recall attempts, but if the delay between recall attempts is long, forgetting can set in and errors tend to increase across recall trials (see also Odinot, Wolters, & Lavender, 2009; Turtle & Yuille, 1994).

In addition to the delay between recall attempts, the nature of the study material can also contribute to whether repeated retrieval would benefit or harm later recall. When the study materials are designed to elicit false recall, repeated retrieval can impair memory accuracy. For example, repeated retrieval increases false recall of a nonpresented word (e.g., sleep) after subjects have studied its related words (e.g., bed, rest, awake, etc., McDermott, 2006; Payne, Elie, Blackwell, & Neuschatz, 1996). Similarly, repeatedly imagining an action (e.g., picking up a pencil) increases the likelihood that one would mistakenly remember having performed that action (Garry, Manning, Loftus, & Sherman, 1996; Thomas, Bulevich, & Loftus, 2003). Moreover, repeated retrieval increases source monitoring errors for visually confusable materials (such as a lollipop and a magnifying glass, Henkel, 2004). Most relevant to current purposes is the finding that repeated testing can increase the recall of misleading information (Register & Kihlstrom, 1988; Roediger et al., 1996). Similar results have been shown when subjects are repeatedly interviewed in a misleading, suggestive manner (e.g., Melnyk & Bruck, 2004). Clearly, for materials that are designed to evoke erroneous memories, repeated retrieval can have harmful effects.

Unlike materials designed to elicit errors, repeated retrieval of neutral materials typically produces hypermnesia. Hypermnesia refers to a net increase in accurate recall over multiple retrieval attempts (for a review, see Payne, 1987). Due to its legal relevance, researchers have examined the effects of repeated retrieval in the realm of eyewitness memory. Here, too, hypermnesia is regularly found, and it occurs for a variety of materials, including videos (Otani et al., 2008; Scrivner & Safer, 1988), violent events (Bornstein et al., 1998), and autobiographical events (Bluck et al., 1999; see also Hershkowitz & Terner, 2007, where reminiscence, but not hypermnesia, is found for children). In addition to producing hypermnesia, repeated retrieval can slow forgetting of the studied material (Roediger & Karpicke, 2006). The impressive benefits of retrieval on long-term retention have been found across a wide variety of materials (Carpenter, Pashler, Wixted, & Vul, 2008; Chan, 2010; Wheeler, Ewers, & Buonanno, 2003).

In sum, repeated retrieval can benefit memory performance in the short-term (i.e., hypermnesia across recall trials) and in the long-term (i.e., a slowing of forgetting) when it occurs in a nonleading, neutral environment. However, when repeated retrieval occurs **following** the introduction of misleading information, or when the repeated questioning happens in a suggestive manner, it can increase errors in memory report. In the current study, the repeated recall attempts occurred **before** misinformation was presented.

The Effects of (Single) Prior Retrieval on Eyewitness Suggestibility

If repeated retrieval can enhance memory performance, can it be implemented to protect eyewitnesses from the debilitating effects of ***misleading postevent information***? Several studies have examined whether testing a witness' memory prior to misinformation exposure can reduce eyewitness suggestibility, and the results so far are mixed. In one of the first attempts to examine the effects of prior recall on subsequent eyewitness suggestibility, Geiselman and colleagues (1986) investigated whether the cognitive interview can reduce eyewitness susceptibility to misinformation. It can, at least in comparison to a standard recall task. However, because the researchers did not include a no-initial-recall control condition, it is not possible to determine whether initial recall had actually reduced eyewitness suggestibility relative to no initial recall (see also Holliday, 2003; Memon, Holley, Wark, Bull, & Kohnken, 1996). Recently, Memon and colleagues (2010) examined whether recalling a witnessed event can reduce subsequent eyewitness suggestibility. In this experiment, participants completed either a free recall test or the cognitive interview before or after misinformation exposure in a factorial design. Participants generated their own misinformation during a forced fabrication phase and took a yes/no recognition test one week later. Memon et al. found that false recognition was reduced when the cognitive interview was given before, rather than after, misinformation fabrication. Notably, however, initial free recall produced no such benefit.

Of most relevance to the current study is a series of experiments reported by Chan and colleagues (Chan & Langley, 2011; Chan et al., 2009; Thomas, Bulevich, & Chan, 2010). In these experiments, participants first viewed a video event and then either took a cued recall test or performed a distractor task. Afterwards, participants were exposed to misinformation via a postevent narrative that purportedly recapped the video, which was followed by a final memory test. Chan et al. reported a surprising finding – initial testing (relative to a distractor task) increased, rather than decreased, participants' susceptibility to later misinformation. In a subsequent study, Chan and Langley showed that this ***retrieval-enhanced suggestibility (RES)*** effect is long-lasting and can occur regardless of whether misinformation is presented immediately or one week following the initial test. Thomas et al. further extended this RES effect to a multiple choice recognition test and showed that participants are highly confident when they report misinformation in the RES procedure. However, they also reported a boundary condition for RES – it is eliminated when participants are given a warning about the accuracy of the postevent narrative (see also Lane, Mather, Villa, & Morita, 2001, who reported no RES effect with a brief, 5-min free recall initial test and a source monitoring final test).

To account for this RES effect, Chan et al. (2009) suggested that the initial test questions might have inadvertently 'highlighted' some aspects of the witnessed event. When participants encounter new information (or misinformation) regarding these details in the postevent narrative, the new information captures participants' attention and becomes particularly well encoded. This enhanced learning of the misinformation is then revealed as RES. This hypothesis has received additional support in subsequent reports (Chan & Langley, 2011; Thomas et al., 2010) and can account for the finding reported by Memon et al. (2010). Specifically, because participants generated their own misinformation, the encoding of such misinformation was unlikely to be enhanced by the initial recall test. As a result, no RES effect would be expected (see too Pansky & Tenenboim, 2011). We return to a more thorough exploration of this hypothesis and its implications in the General Discussion.

The Effects of Repeated Retrieval on Eyewitness Suggestibility

Although several studies have examined the effects of initial recall on subsequent eyewitness suggestibility, no research to date has examined the effects of taking ***repeated initial recall tests*** on later eyewitness suggestibility. Based on previous research, two predictions can be drawn. First, given that taking an initial recall test can lead to increased susceptibility to misinformation, taking multiple initial tests may further exacerbate eyewitness suggestibility.

That is, the repeated recall tests may cause the later misinformation to capture participants' attention even more during the postevent narrative (relative to a single test), thus enhancing the learning of these misleading details. Alternatively, taking repeated recall tests may ***reduce*** subsequent eyewitness suggestibility and eliminate RES. This prediction can be drawn because of two reasons. First, repeated retrieval is a powerful memory enhancer, and eyewitnesses are more likely to resist misinformation when the original memory is stronger (Loftus, 1979; Marche, 1999; Peterson, Parsons, & Dean, 2004); therefore, recalling details of an event multiple times may lead to a reduction in later suggestibility. Second, one component of the cognitive interview involves repeated retrieval. Provided that the cognitive interview may be effective at reducing subsequent eyewitness suggestibility (Memon et al., 2010), repeated retrieval may show similar benefits. Essentially, this prediction suggests that the relation between repeated testing and suggestibility is nonlinear (for a recent demonstration of a quadratic relation between repeated questioning and children's witness memory accuracy, see Krahenbuhl, Blades, & Eiser, 2009). Note that in addition to repeated retrieval, the cognitive interview uses other techniques that are absent from the current study (e.g., rapport building, context reinstatement, free recall followed by cued recall, and recall from multiple perspectives and in various temporal orders). Therefore, even if repeated retrieval alone cannot suppress the misinformation effect in the present study, it remains possible that the cognitive interview can.

Overall Design and Materials of the Present Experiments

To provide a relatively comprehensive investigation of the effects of repeated retrieval on eyewitness suggestibility, we examined three variables across four experiments. The first variable was the number of initial tests taken (0, 1, 3, 5, and 6 across the various experiments). The second was the delay that separated the initial and final tests (i.e., 30 min or 1 week). The third was whether the testing manipulation (i.e., nontested vs. tested) occurred between- or within-subjects. This is the first time the RES procedure has been implemented in both a between- and within-subjects design. This manipulation served two purposes. First, given its practical implications, it is important to examine the generality of the RES phenomenon across different experimental designs. Indeed, the between-subjects and within-subjects comparisons ask different questions regarding RES. In a between-subjects comparison, the question is whether testing increases eyewitness suggestibility for a particular witness. In a within-subjects comparison, the question is whether a tested item is more likely to be altered by misinformation than a nontested item. Second, due to the influence of retrieval on subsequent memory of initially nonretrieved materials (e.g., Anderson, 2003; Chan, McDermott, & Roediger, 2006), it is possible that RES may not occur in a within-subjects design. Several studies have indeed shown that recalling part of an event can impair subsequent memory of the remainder of the event (e.g., Garcia-Bajos, Migueles, & Anderson, 2009; MacLeod, 2002; Shaw, Bjork, & Handal, 1995). We expand on this phenomenon during the Introduction to Experiment 2.

All four experiments used the same general design. First, participants watched an episode of the television program "24." They then completed one cued recall test, multiple recall tests, or an unrelated task. If testing was manipulated within-subjects (as in Experiments 2 and 4), then all participants received the initial test(s), but the initial test(s) included only half of the final test items. After a filled retention interval (i.e., 30 min or 1 week, depending on the experiment), participants listened to an audio narrative that contained the misleading information, which was followed by the final test.

In every experiment, we varied the number of initial tests taken and the item type on a given test (i.e., consistent, control, misleading). The item type variable was manipulated during the postevent narrative. For example, if the original video depicted a character driving an SUV, the narrative might report that the character drove an SUV (i.e., ***consistent***), that the character drove (i.e., ***control***, with no information on the type of vehicle), or that the character drove a pickup truck (i.e., ***misleading***). All questions were answerable, with the misinformation involving the substitution of a correct detail with an incorrect one. Delay was varied across experiments. Specifically, the delay between the initial and final tests was 30 min in Experiments 1 and 2, and 1 week in Experiments 3 and 4. Lastly, the within- vs. between-subjects design was also manipulated across experiments. Experiments 1 and 3 manipulated the test vs. no-test variable between-subjects, whereas Experiments 2 and 4 manipulated the same variable within-subjects. To compensate for the reduction of items when using a within-subjects design relative to a between-subjects design, Experiments 2 and 4 used 36 items whereas Experiments 1 and 3 used a set of randomly-selected 24 items (see Appendix A). All participants in the current experiments were undergraduate students from Iowa State University who participated for research credits. Moreover, all participants were permitted to participate in only one experiment.[1]

Experiment 1

Method

Participants. There were 122 participants in Experiment 1. Fourteen participants were excluded from analyses for reasons including lack of proficiency in English, failure to follow instructions, and an experimenter error. Therefore, all analyses were based on data provided by the remaining 108 participants. No participant had seen the video event within four months prior to participation.

Design. Experiment 1 used a 3 (item type: consistent, control, misleading) X 3 (initial tests: 0, 1, 5) mixed design. Item type was manipulated within-subjects and the number of tests was manipulated between-subjects. Thirty-six participants occupied each between-subjects condition.

Materials and Procedure. Participants completed the study in groups of no more than eight on computer terminals separated by dividers. The experiment started with the witnessed event video that showed a government agent working to stop the assassination of a presidential candidate. Participants were given intentional encoding instructions. The video lasted ~40 min. Following the video, participants in the 1-test and 5-test conditions completed one and five (identical) cued recall tests, respectively. Participants were told that they might be asked the same question repeatedly. They were told to be as accurate as possible and were explicitly given the option to respond with “I don’t know.” No breaks were inserted between each initial test, which contained 24 questions (see the Appendix). Participants had 15 s to type in a response to each question. After the immediate test, participants in all conditions watched a 22 min video as a distractor activity, which depicted the British intelligence agency’s attempt to stop an anti-abortion activist from killing abortion doctors and their families.

Following the distractor video, participants listened to a ~7 min audio narrative. They were told that the narrative recapped the video event. Embedded in the narrative were the 24 critical details that had been tested during the immediate test, eight of these were consistent, eight were control, and eight were misleading. Item type assignment was counterbalanced across participants. After listening to the audio narrative, participants played the videogame Tetris for 5 min as a distractor task. They then completed the final cued-recall test, which contained the same 24 questions as the initial test. The instructions of the final test emphasized that participants should answer the questions based on their memory of the video event and did not mention the audio narrative. Participants were given 15 s to respond to each question.

Our procedure favored performance for participants in the no initial test condition because the delay between the video and the final test was the shortest in this condition (~34 min), next was the 1-test condition (~40 min), and the delay was the longest in the 5-test condition (~66 min). Therefore, if anything, this design should weaken the beneficial influence of repeated testing on final test performance.

Results and Discussion

Responses were classified into four categories: ***Correct, Misinformation, No Answer or Other Intrusions***. The ***Correct*** and ***Misinformation*** categories are self-explanatory. "I don't know" or unanswered questions were scored as ***No Answer***, and the ***Other Intrusions*** category referred to reports that matched neither the original nor misleading detail. The alpha level was set at .05. Partial eta squared (η_p^2) indicates effect size for analysis of variance (ANOVA), and Cohen's *d* indicates effect size for t-tests. We first report results from the initial test, particularly those regarding hypermnesia; we then report results from the final test.

Initial Test. Across the 1-test and 5-test conditions, .52 of the initial test responses were correct, .04 spontaneously matched the misinformation given later, .11 were classified as No Answer, and the remaining .34 were classified as Other Intrusions. See Table 1 for correct recall probabilities as a function of initial test trial. Overall, initial test accuracy did not differ between the 1-test and 5-test conditions, $t < 1$ ($M = .51$ and $M = .53$, respectively). However, participants in the 5-test condition demonstrated significant hypermnesia. That is, their recall probability rose over the five initial test trials, $F(4, 140) = 10.68$, $p < .01$, $\eta_p^2 = .23$.

Final Test. Separate analyses were conducted for misinformation recall and accurate recall probabilities (see Table 2 for means). Most importantly, ***misinformation recall probability*** increased with the number of initial tests taken, $F(2, 105) = 10.04$, $p < .01$, $\eta_p^2 = .16$. Planned comparisons showed that both participants in the 1-test ($M = .48$) and 5-test ($M = .55$) conditions recalled more misinformation than those in the no-test condition ($M = .29$), both *t*s > 3.35, *p*s < .01, *d*s > .79. However, the misinformation recall probabilities did not differ significantly between the 1-test and the 5-test conditions, $t < 1$, possibly due to low statistical power (observed power = .14).

A 3 (initial tests: 0, 1, 5) X 3 (postevent information: consistent, control, misleading) ANOVA was conducted for ***correct recall probability***. The main effect for postevent information and the interaction were significant, *F*s > 6.73, $p < .01$, $\eta_p^2 > .11$, but the main effect for number of initial tests was not, $F < 1$. Planned comparisons show that correct recall probability increased with testing for consistent items, $F(2, 105) = 4.93$, $p = .01$, $\eta_p^2 = .09$, but not for control items, $F(2, 105) = 1.10$, $p = .34$, though the latter's correct recall probabilities did rise numerically (from .47 to .53) with the number of initial tests taken. Importantly, for both the consistent and control items, the highest accurate recall probability was obtained by participants in the 5-test condition, showing that repeated retrieval can enhance eyewitness memory for neutral materials. Remarkably, this testing effect was reversed when one encountered misinformation, $F(2, 105) = 2.40$, $p = .10$, $\eta_p^2 = .04$.

To further explore the influence of testing on eyewitness suggestibility, we computed false recall probabilities based on output-bound measures. In the laboratory, researchers are aware of the information presented to participants and can produce input-bound measures, which are based on, for example, the number of correct responses out of the total number of items presented during encoding. However, in actual criminal cases, output-bound measures (that is, measures that are based on the total responses volunteered at output, removing omissions) might be equally important (Fisher, 1996; Koriat & Goldsmith, 1996a; but see Bjork & Wickens, 1996, for caution against using such conditional measurements of accuracy). We therefore examined the effects of initial testing on output-bound misinformation susceptibility. Similar to the input-bound measures, the output-bound misinformation recall probability increased with number of initial tests taken, $F(2, 105) = 7.97$, $p < .01$, $\eta_p^2 = .13$, going from .32 with no initial test to .50 with one initial test and .57 with five initial tests.

Experiment 2

In Experiment 2, we examined the influence of repeated testing on eyewitness suggestibility in a within-subjects design. To that end, all participants completed an initial test phase, but the initial test(s) included only half of the items that were to appear in the postevent narrative and the final test, and misinformation susceptibility was compared between the nontested and the tested items (i.e., the RES comparison). Number of initial tests was manipulated between-subjects, such that some participants received only one initial test, some received three initial tests, and others received six initial tests.

Examining RES in a within-subjects design is important from an applied perspective. In actual criminal cases, most witnesses must repeatedly recall details of a witnessed event prior to providing testimony in court, and it is likely that questioning will become more comprehensive as the investigation progresses. That is, a witness may need to answer new questions later in the investigation that he/she had not been asked earlier. It is therefore important to examine whether the RES effect would generalize to a within-subjects situations. Instead of asking whether witnesses who had previously recalled an event would be more suggestible than witnesses who had not performed any previous tests (as in Experiment 1), Experiment 2 addresses whether an eyewitness would be more susceptible to misinformation for a previously tested item than a nontested item. From the perspective of psychology and the law, this is a very different, yet equally important, question.

Aside from its applied implications, it is also interesting to examine RES in a within-subjects design for theoretical reasons. Specifically, we expect the RES effect to be smaller in magnitude in a within-subjects design. We arrive at this prediction based on two related findings in the literature – retrieval-induced forgetting (for a recent review, see Anderson & Levy, 2009) and retrieval-induced facilitation (Chan et al., 2006). Retrieval-induced forgetting refers to the finding that recalling a subset of studied information can impair subsequent retrieval of the nontested but related information. Retrieval-induced facilitation refers to the opposite finding; that is, recalling some studied information can enhance later recall of the nontested information. The apparent contradiction between these findings has been addressed in detail elsewhere (Chan, 2009). For current purposes, what is important is that both of these findings would lead one to predict that RES would diminish in a within-subjects (relative to between-subjects) design. If retrieval-induced forgetting were to occur in the current context, then the initial test would weaken subsequent memory of the nontested items, thus increasing the likelihood that these items would be altered by misinformation later (Saunders & MacLeod, 2002, but see also Odinot et al., 2009). Conversely, if repeated testing causes covert retrieval of the nontested items (i.e., retrieval-induced facilitation, cf. Garcia-Bajos et al., 2009), then these nontested items would become more similar to the tested items, thus reducing the difference in suggestibility between the tested and nontested items.

Method

Participants. There were 114 participants in Experiment 2. Four participants were excluded from analyses because English was not their primary language. Among the remaining 110 participants, 36 were in the 1-test condition, 38 were in the 3-test condition, and 36 were in the 6-test condition.

Design. The experiment used a 2 (testing condition: tested vs. nontested) X 3 (tests: one, three, six) X 3 (item type: consistent, control, misleading) mixed design. Testing condition and item type were manipulated within-subjects; number of tests was manipulated between-subjects.

Materials and Procedure. The materials and procedure were identical to those in Experiment 1 with the following exception: Because the testing condition (i.e., tested vs. nontested) was manipulated within-subjects, all participants received the immediate test. There were 18 questions on the immediate test and 36 questions on the final test. Whether an item was included on the initial test was counterbalanced across participants. Participants took the immediate test either once, three times, or six times. The narrative included 36 critical details, of which 12 were consistent, 12 were control, and 12 were misleading. These critical details included those used in Experiment 1 as well as 12 additional details. None of our conclusions changed regardless of whether we analyzed the data based on all 36 questions or only the 24 questions that appeared in all four experiments.

Results and Discussion

Initial Test. Overall, .58 of participants' responses were correct, .05 matched the misinformation, .30 were Other Intrusions, and the remaining .08 were No Answer. Similar to Experiment 1, repeated recall led to hypermnesia in both the 3-test and 6-test conditions (see Table 1 for means), $Fs > 2.42$, $ps < .05$, $\eta_p^2s > .07$.

Final Test. A 2 (testing condition: tested, nontested) X 3 (number of initial tests: 1, 3, 6) ANOVA was conducted to examine the effects of initial testing on ***misinformation recall*** probability (see Table 3). The main effect of testing condition indicates that a significant RES effect was found. That is, participants were more likely to report misinformation for a tested item (M = .48) than for a nontested item (M = .35), $F(1,108) = 27.45$, $p < .01$, $\eta_p^2 = .20$. Moreover, the interaction shows that the magnitude of RES increased with repeated initial tests, $F(2, 108) = 6.34$, $p = .01$, $\eta_p^2 = .11$, such that the RES effect increased from a nonsignificant .05 with a single initial test, $t(35) = 1.25$, $p = .22$, $d = .18$, to .09 with three initial tests, $t(37) = 2.06$, $p = .05$, $d = .35$, and .26 with six initial tests, $t(35) = 5.38$, $p < .01$, $d = .91$. Similar to Experiment 1, the output-bound misinformation probabilities provide conclusions consistent with those from the input-bound measures. The output-bound RES effect rose from .04 with one initial test to .07 with three initial tests to .24 with six initial tests, interaction $F(2, 108) = 5.56$, $p = .01$, $\eta_p^2 = .09$. These findings illustrate the harmful effects of repeated testing on eyewitness suggestibility.

To examine the effects of repeated testing on ***accurate recall***, we conducted separate ANOVAs for each type of postevent information (i.e., consistent, control, misleading). A 2 (testing condition: tested, nontested) X 3 (number of initial tests) ANOVA showed that testing enhanced accurate recall of both the consistent items, $F(1, 108) = 29.57$, $p < .01$, $\eta_p^2 = .20$, and the control items, $F(1, 108) = 3.31$, $p = .07$, $\eta_p^2 = .03$. However, increasing the number of initial tests did not alter the magnitude of these testing effects, $F < 1$. The misleading items, however, did not show a testing benefit, $F(1, 108) = 1.16$, $p = .21$. In fact, a reversed testing effect became evident as the number of initial tests increased, interaction $F(2, 108) = 3.30$, $p = .04$, $\eta_p^2 = .06$.

Experiments 1 and 2 show that repeated testing can increase eyewitness suggestibility to subsequent misinformation. However, perhaps due to the short delay between the initial and final tests, we failed to find a robust testing benefit for the control items. Would repeated testing continue to increase eyewitness suggestibility if its memory enhancing effects can be revealed on the control items? Recently, Chan and Langley (2011) reported that testing can increase eyewitness suggestibility even after a one-week delay, where significant testing benefits were observed for the control items, but they did not examine the influence of ***repeated initial tests*** on subsequent eyewitness suggestibility. Given that the benefits of repeated testing (over taking a single initial test) is particularly pronounced after a long retention interval (Karpicke & Roediger, 2007), and that actual eyewitnesses often provide testimony after a lengthy delay, we examined the long-term effects of repeated testing on eyewitness suggestibility in Experiments 3 and 4. To that end, a one-week interval was inserted between the initial test and the final test. To facilitate comparison across experiments, Experiment 3 was modeled after Experiment 1 (manipulating testing between-subjects) and Experiment 4 was modeled after Experiment 2 (manipulating testing within-subjects).

Experiment 3

Method

Participants. There were 156 participants in Experiment 3. Data from 55 participants were excluded from analyses, because 32 failed to attend the second testing session, and 23 were excluded based on reasons similar to those in Experiment 1. Therefore, all analyses were based on the remaining 101 participants; with 34 in the no-test condition, 34 in the 1-test condition, and 33 in the 5-test condition. Attrition between the experimental sessions did not produce a subject selection artifact, because the initial test accuracy did not differ between the dropout and non-dropout participants, $t(81) = 1.63$, $p = .11$.

Materials and Procedure. The materials and procedure of Experiment 3 were identical to those of Experiment 1 except that the retention interval was changed to one week. This delay replaced the distractor video. That is, subjects watched the critical event video, completed the initial test, and were then dismissed for one week. Upon their return, they listened to the audio narrative, played Tetris, and then completed the final test.

Results and Discussion

Initial Test. Across the 1-test and 5-test conditions, .53 of the initial test responses were correct, .03 matched the misinformation, .13 were No Answers, and .30 were classified as Others Intrusions. There was no difference between the 1-test and 5-test conditions in overall initial test accuracy ($M = .51$ and $M = .55$, respectively), $t(65) = 1.29$, $p = .20$, but a significant hypermnesia effect was found for participants in the 5-test condition (see Table 1 for means), $F(4, 128) = 9.54$, $p < .01$, $\eta_p^2 = .23$.

Final Test. Consistent with the results from Experiments 1 and 2, ***misinformation recall*** probability rose with the number of initial tests taken, $F(2, 98) = 4.28$, $p = .02$, $\eta_p^2 = .08$ (see Table 4). That is, taking a single initial test increased (though not significantly) misinformation recall probability from .44 to .51, $t(66) = 1.50$, $p = .14$, $d = .36$, and taking five initial tests further increased misinformation recall probability to .59, $t(65) = 2.98$, $p < .01$, $d = .73$ (compared to no initial testing).

To examine the effects of initial testing on ***accurate recall***, separate one-way ANOVAs were conducted for each item type with number of initial tests as the independent variable. As expected, a powerful testing effect was found for the consistent items and the control items, *F*s >

14.08, *p*s < .01, η_p^2s > .22. In fact, the control items showed a clear advantage of taking multiple initial tests on final recall (*M* = .55) relative to a single initial test (*M* = .46), *t*(65) = 2.28, *p* = .03, *d* = .56. Remarkably, despite the significant RES effect on misinformation recall, the misleading items showed a testing effect on accurate recall, *F*(2, 98) = 7.54, *p* < .01, η_p^2 = .13, and this testing effect was found for both the 1-test and 5-test participants, both *t*s > 3.61, *p*s < .01, *d* > .88. Two possibilities are consistent with this finding. The first is that initial testing enhanced both retention of the witnessed event (leading to an increase in accurate recall) and the learning of the misinformation. The second possibility is that initial testing relaxed participants' response criterion on the final test. Although our results are consistent with both possibilities, it seems unlikely that initial testing would lead to a criterion shift for two reasons. First, to our knowledge, no existing data would prompt a criterion shift prediction a priori. Second, we found no such pattern in Experiments 1 and 2 (see Table 2 and Table 3).

Unlike Experiments 1 and 2, the output-bound misinformation recall probabilities here were not affected by initial testing (.54 for no-test, .55 for 1-test, and .60 for 5-test), *F* < 1. Why did the output-bound data produce a different pattern than the input-bound data? From a quantitative perspective, the only difference between input- and output-bound measures is that the former include omission (which is the No-Answer) responses in the denominator and the latter do not. Note that No-Answer responses accounted for a very small proportion of the total responses in Experiment 1 (*M* = .05) and Experiment 2 (*M* = .04), so the input- and output-bound data should resemble each other. However, a one-week retention interval was used in Experiment 3, thus allowing significant forgetting to occur. Because we did not use a forced report procedure, participants were therefore free to omit any responses that they deem unlikely to be correct, and participants who did not take an initial test omitted .16 of the questions.

However, because testing slowed forgetting, participants in the 1-test and 5-test conditions omitted only .06 and .01 of the questions, respectively.

Because output-bound measures disregard the No-Answer responses from the denominator, response probabilities in conditions that contain a large proportion of No-Answer responses would rise accordingly. For example, the raw misinformation recall probability in the no-test condition was .44, when one divide this probability by the total output minus No-Answers (.84), the output-bound misinformation recall probability rose to .52. However, because participants in the 1-test and 5-test conditions rarely used the No-Answer option, the probability of misinformation recall had virtually no room to move up. Therefore, output bound measures, under some conditions, can mask the effects of variables that typically influence forgetting (e.g., Fisher, 1996, reported that output bound accuracy showed little forgetting over long periods of delay). We return to a more thorough discussion of this idea in the General Discussion.

In sum, Experiment 3 revealed that repeated testing can increase both accurate recall and misinformation susceptibility following a one-week retention interval. In contrast to Experiment 1, it is clear that the delay has allowed the advantage of testing to emerge for the consistent and control items. Perhaps more surprisingly, the misleading items also demonstrated a testing benefit on its accurate recall probability, even though repeated testing also increased its misinformation recall probability. In Experiment 4, we examined the effects of repeated testing on delayed eyewitness suggestibility in a within-subjects design. Given that the RES effect is smaller in a within-subjects design (see Experiment 2) and the testing benefit is particularly powerful after a delay, it is possible that they would combine to eliminate RES altogether (cf., Pansky & Tenenboim, 2011). In addition, we included a modified-modified free recall (MMFR, Barnes & Underwood, 1959) test following the regular "final test" in this experiment. The MMFR test was designed to reveal the accessibility of the original event memories and the misinformation, while minimizing the influence of response competition and metamemorial control processes that lead to response withholding (Higham, 2002; Roebers & Fernandez, 2002).

Experiment 4

Method

Participants. One-hundred-sixty-one students participated in this experiment. Data from 48 participants were excluded because 38 failed to attend the second testing session and 10 for reasons similar to those in Experiment 1. All analyses were based on the remaining 113 participants; with 38 in the 1-test condition, 35 in the 3-test condition, and 40 in the 6-test condition. No subject selection problem was found when we compared the initial test accuracy between participants who attended the second session and those who did not, $t < 1$.

Materials and Procedure. The materials and procedure of Experiment 4 were identical to those of Experiment 2 except that the retention interval following the initial tests was changed to one week, and an MMFR test was administered. The MMFR test contained the same questions as the final test, but participants were given 25 s to respond to each question and were instructed to recall details from both the video and the audio narrative even if the details contradicted each other. Participants were not asked to specify the source of their recalled details, and their responses were scored in the same way as in the regular test. Note that because the MMFR test was designed to elicit multiple responses for the misleading questions, the combined probabilities of Correct, Misinformation, Other, and No Answer could, and often did, exceed 1.

Results and Discussion

Initial Test. Overall, .58 of the responses were correct, .04 matched the misinformation, .31 were Other Intrusions and .08 were No Answer. No difference was detected in accurate recall probability across the 1-test, 3-test, and 6-test conditions ($M = .59$, $M = .55$, $M = .58$, respectively), $F < 1$. In addition, the hypermnesia effect was significant in the 6-test condition (see Table 1 for means), $F(5, 195) = 7.25$, $p < .01$, $\eta_p^2 = .16$, but not in the 3-test condition, $F(2, 68) = 2.26$, $p = .11$, $\eta_p^2 = .06$.

Final Test. Recall probability of ***misinformation*** rose with the number of initial tests taken, $F(2, 110) = 2.60$, $p =.08$, $\eta_p^2 = .05$ (see Table 5). However, unlike the prior three experiments, no RES effect was found, $F < 1$. Moreover, there was no interaction between testing condition and number of initial tests, $F(2, 110) = 1.21$, $p = .30$. That is, although repeated testing increased misinformation recall probabilities, this increase was comparable between the nontested and tested items, such that no significant RES effect was observed across all testing conditions, all *t*s < 1.44, $p > .15$. Likewise, output-bound measures also revealed no significant RES effect, all *t*s < 1.13, $p > .27$.

We now examine the effects of repeated testing on ***accurate recall***. For consistent items, there was a significant testing effect, $F(1, 111) = 49.43$, $p < .01$, $\eta_p^2 = .31$, which increased with the number of initial tests taken, $F(2, 110) = 4.53$, $p = .01$, $\eta_p^2 = .08$. Specifically, the testing effect increased from .08 with a single initial test, $t(37) = 1.78$, $p = .08$, $d = .31$, to .25 and .19 with three and six initial tests, respectively, *t*s > 4.03, *p*s < .01, *d*s > .80. For the control items, there was a substantial (.26) testing benefit, $F(2, 110) = 108.26$, $p < .01$, $\eta_p^2 = .50$, but this testing effect did not vary with number of initial tests, $F(2, 110) = 2.06$, $p = .13$, $\eta_p^2 = .04$, $p = .13$. For the misleading items, similar to Experiment 3, a powerful (.18) testing effect is observed, $F(1, 110) = 55.13$, $p < .01$, $\eta_p^2 = .33$. This finding suggests that repeated testing, combined with a long delay, can reduce eyewitness suggestibility in a within-subjects context. We now present data from the MMFR test.

MMFR Test. Separate 2 (testing condition: tested, nontested) X 3 (number of initial tests: 1, 3, 6) ANOVAs examined the effects of testing on ***accurate recall*** of each item type in the MMFR test (see Supplementary Materials for the full data set). Similar to results from the regular cued recall test, a testing benefit was found for all item types, all *F*s > 32.30, *p*s < .01, η_p^2 > .22. No other effects were significant.

A 2 (testing condition) X 3 (number of initial tests) ANOVA was conducted to examine whether repeated testing enhanced misinformation learning, despite there being no RES effect in the regular cued recall test. Unlike the data from the cued recall test, initial testing increased misinformation recall probability in the MMFR test, $F(1, 110) = 6.28$, $p = .01$, $\eta_p^2 = .05$. Specifically, participants were more likely to report the misinformation for the tested items (*M* = .61) than the nontested items (*M* = .54). Planned comparisons revealed a marginally significant RES effect for both the 3-test (a .09 difference) and the 6-test (a .10 difference) conditions, $t(34) = 1.79$, $p = .08$, $d = .37$, and $t(39) = 1.96$, $p = .06$, $d = .35$, respectively. The RES effect for the 1-test condition was not significant but was also in the predicted direction (a .04 difference), $t < 1$.

Why did an RES effect occur in the MMFR test but not in the cued recall test? In the cued recall test, participants were able to exert control over the reporting of information (Koriat & Goldsmith, 1996b; Koriat, Goldsmith, & Pansky, 2000). When participants encountered a question with multiple possible answers (e.g., the correct item and the misinformation), they must choose one to report, and this decision process should depend on the relative strength of the two items in memory (Thomas et al., 2010). The results of the MMFR test indicate that testing had a powerful strengthening effect on the memory of the original item. Collapsed across all testing conditions, the ***accurate recall*** probability of the misleading items was .30 for the nontested items and .54 for the tested items. Although testing did enhance learning of the ***misinformation*** (.54 for nontested vs. .61 for tested), this increase was small relative to the benefit for the original item. We believe that this differential strengthening of the original and misleading items contributed to the elimination of the RES effect in cued recall.

General Discussion

Four major findings emerged in the present study. First, taking multiple initial tests can increase eyewitness suggestibility. Second, the RES effect can happen on an item level (i.e., in a within-subjects situation) as well as on a participant level (between-subjects). Third, the size of the RES effect is smaller on an item level than on a participant level. Fourth, the item-level (but not the participant-level) RES effect can be eliminated when a long delay separates the initial and final test. We now discuss the theoretical and applied implications of these findings.

Repeated Testing Strengthens Retrieval-Enhanced Suggestibility

In all four experiments, there was one clear, replicable pattern – taking repeated initial tests increased eyewitness suggestibility to later misinformation. Across these four experiments, the average effect size of RES (as measured in Cohen's *d)* with a single initial test was .30, but it was .76 with five or six initial tests. Even armed with the knowledge of RES, this finding is surprising. How can misinformation alter memory reports more effectively after multiple recall attempts than after one, or even no, recall attempt? The data from the MMFR test in Experiment 4 can shed light on this question. Similar to the explanation offered previously (Chan & Langley, 2011; Thomas et al., 2010), the RES effect might occur because initial testing inadvertently drew attention to a subset of the items in the witnessed event. When new information regarding these items is presented, it captures participants' attention. Following multiple tests (as opposed to one test), the tested details become even more memorable, and misinformation regarding these details further captures participants' attention when it is presented. This situation somewhat resembles that of hypercorrection, where high-confidence errors are more readily corrected by feedback than low-confidence errors (Butterfield & Metcalfe, 2001; Kulhavy, Yekovich, & Dyer, 1976). In a typical hypercorrection experiment, participants answer general knowledge questions that are likely to elicit errors (e.g., What is the capital of Australia? Answer: Canberra), provide confidence for their answers, and then receive the correct answer as feedback. Surprisingly, errors made with high confidence are more often corrected than those made with low confidence.

Applying this logic to the present context, it is possible that some participants have assumed that all information presented in the postevent narrative was accurate, thus treating the narrative as a form of feedback. If this is the case, then manipulations that increase the confidence of the original event details should increase their susceptibility to later misinformation, and repeated testing is one such manipulation. Indeed, testing has been demonstrated to increase the confidence associated with a retrieved memory (Shaw, 1996; Thomas et al., 2010), and repeated testing can further inflate the confidence associated with these memories (Shaw & McClure, 1996). When new information regarding these memories is presented, participants might experience a feeling of surprise (for similar arguments, see Fazio & Marsh, 2009); if participants fail to question the accuracy of this new information, then they might "hypercorrect" their original responses. Consistent with this notion, the RES effect is eliminated when participants are told explicitly that information presented in the postevent narrative might be incorrect (Thomas et al.).

Can Initial Testing Help, Rather Than Hurt, Eyewitness Memory Accuracy?

Perhaps the most encouraging finding in the current study is that taking a single initial test (rather than multiple initial tests) can, under some circumstances (see Experiment 4), reduce eyewitness suggestibility. To consider the implications of this finding, we begin by examining the impact of repeated testing on performance during the initial test phase. Strikingly, across all four experiments, the greatest hypermnesia occurred during the second initial test (see Table 1). Indeed, performance tended to peak at the second test trial, and further testing did not confer additional benefits. Moreover, taking a single initial test already produced considerable benefits for delayed recall. Across the four experiments, the effect size (in Cohen's *d*) of taking one initial test on final accurate recall was .60, .83, and .35 for the consistent, control, and misleading items, respectively. With five or six initial tests, these effect sizes became .81, .99, and .20. Therefore, taking five or six initial tests led to a modest benefit on delayed accurate recall compared with taking only one test for the consistent and control items, but this advantage was more than offset by a substantial increase in participants' suggestibility (from .30 to .76, as mentioned previously).

Although providing a warning can minimize RES (Thomas et al., 2010), one concern with such an approach is that warnings might be more effective in the laboratory than in the field. Specifically, in laboratory settings, targeted warnings can be delivered because the experimenter is aware of the source of the misinformation (e.g., the experimenter might inform the participant that the **postevent narrative** might contain misinformation). However, in actual criminal investigations, it is unclear whether or from where a witness might have encountered misinformation. Consequently, warnings are likely less specific in the field than in experiments (for a similar argument on warning in the context of post-identification feedback, see Lampinen, Scott, Pratt, Leding, & Arnal, 2007), and the specificity of a warning can have a large impact on its effectiveness. For example, general warnings that simply question the credibility of the postevent narrative (e.g., Greene, Flynn, & Loftus, 1982) often produce weaker effects than specific warnings that directly inform participants not to report information from the narrative (e.g., Lindsay, 1990; Wright, 1993). Therefore, “realistic” warnings might not be as effective as those in Thomas et al. It is thus imperative to explore alternative methods that can reduce the impact of RES.

In the current experiments, we found that testing, in combination with a long delay, can eliminate RES on the item (or within-subjects) level but not on the participant (or between-subjects) level (see Pansky & Tenenboim, 2011, for a recent demonstration of a similar result). Before we discuss the implications of this finding, it is important to first examine the two factors that contribute to this effect. The first is that the memorial benefit of testing increases with delay. The second is that the item-based RES effect appears to be weaker than the participant-based RES effect. The former finding is well known, as many researchers have found that testing can slow forgetting, particularly after longer retention intervals (for recent discussions, see Carpenter et al., 2008; Chan, 2010). However, given that a long delay itself does not eliminate RES (as evidenced by the data from Experiment 3 here and also from Chan and Langley, 2011), we must now consider the within- and between-subjects design difference as the crucial contributor to the weaker RES effect in Experiments 2 and 4. As we have mentioned during the Introduction to Experiment 2, two findings suggest that a within-subjects design can reduce the magnitude of RES relative to a between-subjects design. First, the retrieval-induced forgetting argument suggests that initial testing might trigger inhibition of the nontested items, thereby weakening these items and causing them to be more susceptible to misinformation later

(MacLeod, 2002). Alternatively, the retrieval-induced facilitation idea suggests that during the initial test, participants might covertly retrieve target-related memories. For example, when participants attempt to answer a question about the main character's shirt color, they might spontaneously recall the shirt color of his wife, with whom he was conversing. This covert retrieval renders the tested and nontested items more similar, thus weakening the item-based (or within-subjects) RES effect.

Although both retrieval-induced forgetting and facilitation can weaken the item-based RES effect, according to the existing literature, one might expect retrieval-induced facilitation to be the more likely contributor here. Specifically, we used coherent video materials (instead of word lists) in both experiments and a week-long delay to separate the initial and final tests in Experiment 4. These factors can reduce the likelihood of retrieval-induced forgetting (Chan, 2009; Odinot et al., 2009). One way to examine whether retrieval-induced forgetting had occurred in our experiments is to compare the accurate recall probabilities of the nontested control items in the within-subjects experiments (e.g., see the third mean in the top three rows of Table 5) with the control items for the nontested participants in the between-subjects experiments (e.g., see the second mean in the first row of Table 4). If retrieval inhibition had occurred, then the former accurate recall probability would be lower than the latter. Because Experiments 2 and 4 had 36 items whereas Experiments 1 and 3 had only 24 items, we examined the recall probabilities of only the 24 items that appeared in all four experiments to ensure a fair comparison. When considering only these 24 items, results from the first initial test showed that there was no difference in participants' performance across the four experiments (M = .51, .51, .50, .50 for Experiments 1, 2, 3, and 4, respectively) – that is, any difference observed in the final test cannot be attributed to subject differences across experiments. Based on the final test data, it is clear that retrieval-induced forgetting did not occur in our within-subjects experiments, because accurate recall probability of the nontested, control items in Experiment 2 (.49) was not lower than that in Experiment 1 (.47, see first row of Table 1). In fact, the same comparison showed a numerical advantage for Experiment 4 (.25) over Experiment 3 (.19, see first row of Table 3), suggesting that retrieval-induced facilitation might have occurred.

Note that the current experiments were not designed to address the underlying mechanisms that contribute to the difference between the item-based and participant-based RES effect. Rather, these experiments were devised to examine whether RES can occur on an item (or within-subjects) level, and the answer is yes. For present purposes, the important finding is that taking a single initial test can enhance eyewitness memory without increasing suggestibility in a within-subjects context, and this advantage is particularly prominent after a substantial retention interval.

Retrieval-Enhanced Suggestibility or Retrieval-Enhanced Errors?

Although our data generally showed that repeated retrieval can enhance eyewitness suggestibility to misinformation, one issue must be addressed prior to our conclusion section. Specifically, to fully understand the impact of repeated testing on eyewitness memory, one must account for two types of commission (instead of omission) errors – systematic errors produced by exposure to misinformation and nonsystematic errors produced idiosyncratically by the witness (i.e., the “Other Intrusions” as defined here). From a legal perspective, mistaken witness reports are undesirable, regardless of their origin; therefore, it is important to examine whether repeated testing can increase the total errors (i.e., misinformation recall plus other intrusions) produced by eyewitnesses. To simplify comparisons, we compared the total error rate between the no-test and 5/6-test condition across the four experiments. The data are presented in Figure

1. A clear pattern emerged. Repeated testing increased the total errors in recall in Experiments 1 and 2, but it had no influence on total error rates in Experiments 3 and 4. How can one reconcile this discrepancy? The main difference between the first two experiments and the last two experiments was the delay implemented, with the former being 30 min and the latter being one week. It is possible that repeated testing failed to increase the total error rate in the latter experiments because participants were more prone to producing idiosyncratic intrusions when they have forgotten a substantial portion of witnessed event, and some of these intrusions were replaced by misinformation for participants in the repeated testing condition. Hence, initial testing did not alter the overall commission error rate in these experiments. Critically, however, this finding does not change our claim that repeated testing increased eyewitness ***suggestibility to misinformation***.

There is an important distinction between misinformation-based errors and other intrusions, as only the former is systematic and manipulable, and thus much more dangerous in legal settings. For example, a police officer may knowingly introduce misinformation to a suspect in hopes of obtaining a conviction, as reported in some recovered memory cases (Loftus, 1993). Moreover, misinformation errors and other intrusions are likely characterized by different metamemorial experience. Although we did not acquire confidence data in the current experiments, unpublished data collected under similar conditions in our laboratory indicate that misinformation-based errors are accompanied by higher confidence than other intrusions (d = .68, N = 82), which might be crucially important in courtroom situations. In sum, although repeated testing does not always increase the total errors produced by an eyewitness, it does appear that repeated testing can increase eyewitness suggestibility to misleading information.

Concluding Remarks

The finding that repeated recall can exacerbate the misinformation effect has important implications for eyewitness memory. Because repeated questioning is likely necessary in most police interviews (Krahenbuhl, 2007), it is crucial to develop techniques that can reduce its negative impact. Recently, Thomas et al. (2010) reported that people can overcome the debilitating effects of RES when armed with a warning. This is certainly an encouraging finding, though it is unclear whether providing a warning would eliminate RES after ***repeated*** recall. Moreover, as mentioned previously, it might not be possible for criminal investigators to provide the type of precise warnings typically employed in the laboratory. Therefore, it is crucial for future research to investigate methods that can reduce the influence of RES, especially following repeated retrieval attempts.

Although we made an effort to ensure that our methods are somewhat analogous to real-life eyewitness situations (e.g, by using a relatively life-like video event, by testing eyewitness memory over separate sessions), several methodological decisions may limit the generalizability of our results. For example, we used the same video event across all four experiments, we administered the identical recall test in a single experimental session, and the longest retention interval we used was one week. In actual cases, the repeated questions are sometimes administered by different individuals at various intervals and are likely to include different questions. Moreover, the delay involved in actual criminal cases can be far longer than one week. The timing of repeated testing can have a profound effect on memory. In particular, distributed practice has been shown to enhance long-term retention more than massed practice (for a review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Given that increasing the delay between the repeated tests also necessitates an increase in the overall retention interval, it is possible that repeated, distributed tests can reduce the impact of RES. Admittedly, it is impossible for any study to investigate all possible schedules of repeated testing (among other variables), so further research is required to expand on the scope of the present findings. Note, though, for situations where the repeated questions happen in a single interview session, our procedure might have maximized their benefits on memory, because our repeated questions occurred on the computer (a clearly automated system) and participants were told in advance that they might be asked the same questions multiple times. These procedures likely diminished the detrimental effects of repeated questioning relative to how it is conducted in field investigations, where eyewitnesses might feel the social pressure to alter an answer when they were asked the same question again and again by a police officer. Consequently, it is all the more impressive that we found such a powerful RES effect in the current experiments.

In sum, several new findings emerged in the present study. We found that repeated testing can increase accurate recall when the questions are asked in an open-ended, neutral manner. However, when one is given misleading information later, the initial repeated tests can increase, rather than reduce, eyewitness suggestibility. Importantly, the magnitude of this RES effect depends on a number of factors, including whether the testing manipulation occurs within- or between-subjects, the number of initial test taken, and the retention interval. These findings helped to further elucidate the condition under which repeated retrieval would improve, or impair, eyewitness memory.

References

Allen, B.P., & Lindsay, D.S. (1998). Amalgamations of memories: Intrusion of information from one event into reports of another. *Applied Cognitive Psychology, 12*, 277-285.

Anderson, M.C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*, 415-445.

Anderson, M.C., & Levy, B.J. (2009). Suppressing unwanted memories. *Current Directions in Psychological Science, 18*, 189-194.

Ballard, P.B. (1913). Oblivescence and reminiscence. *British Journal of Psychology Monograph Supplements, 1*, 1-82.

Barnes, J.M., & Underwood, B.J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*, 97-105.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*: NY: Macmillan.

Binet, A. (1900). *La suggestibilite*. Paris: Schleicher Freres.

Bjork, R.A., & Wickens, T.D. (1996). Memory, metamemory, and conditional statistics. *Behavioral and Brain Sciences, 19*, 193-194.

Bluck, S., Levine, L.J., & Laulhere, T.M. (1999). Autobiographical remembering and hypermnesia: A comparison of older and younger adults. *Psychology and Aging, 14*, 671-682.

Bornstein, B.H., Liebel, L.M., & Scarberry, N.C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology, 12*, 119-131.

Brown, W. (1923). To what extent is memory measured by a single recall? *Journal of Experimental Psychology, 6*, 377-382.

Butterfield, B., & Metcalfe, J.A. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1491-1494.

Carpenter, S.K., Pashler, H., Wixted, J.T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36*, 438-448.

Cepeda, N.J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354-380.

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

Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory, 18*, 49-57.

Chan, J. C. K., & Langley, M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 248-255.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*, 553-571.

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.

Erdelyi, M.H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6*, 159-171.

Fazio, L.K., & Marsh, E.J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review, 16*, 88-92.

Fisher, R.P. (1996). Implications of output-bound measures for laboratory and field research in memory. *Behavioral and Brain Sciences, 19*, 197.

Garcia-Bajos, E., Migueles, M., & Anderson, M.C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory, 17*, 92-103.

Garry, Maryanne, Manning, Charles G., Loftus, E. F., & Sherman, S.J. (1996). Imagination inflation: Imagining a childhood event inflates confidences that it occurred. *Psychonomic Bulletin & Review, 3*, 208-214.

Geiselman, R.E., Fisher, R.P., Cohen, G., Holland, H., & Surtes, L. (1986). Eyewitness responses to leading and misleading questions under the cognitive interview. *Journal of Police Science and Administration, 14*, 31-39.

Goodman, G.S., & Quas, J.A. (2008). Repeated interviews and children's memory: It's more than just how many. *Current Directions in Psychological Science, 17*, 386-390.

Gray, E. (1993). *Unequal justtice: The prosecution of child sexual abuse*. New York: Free Press.

Greene, E., Flynn, M.S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior, 21*, 207-219.

Henkel, L. A. (2004). Erroneous memories arising from repeated attempts to remember. *Journal of Memory and Language, 50*, 26-46.

Hershkowitz, I., & Terner, A. (2007). The effects of repeated interviewing on children's forensic statements of sexual abuse. *Applied Cognitive Psychology, 21*, 1131-1143.

Higham, P.A. (2002). Strong cues are not necessarily weak: Thomson and tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition, 30*, 67-80.

Holliday, B.E. (2003). The effect of a prior cognitive interview on children's acceptance of misinformation. *Applied Cognitive Psychology, 17*, 443-457.

Hyman, I. E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language, 35*, 101-117.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*, 151-162.

Koriat, A., & Goldsmith, M. (1996a). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences, 19*, 167-228.

Koriat, A., & Goldsmith, M. (1996b). Monitoring and control processes in strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481-537.

Krahenbuhl, S.J. (2007). *The effect of question repetition on young children's eyewitness testimony*. Sheffield: University of Sheffield.

Krahenbuhl, S.J., Blades, M., & Eiser, C. (2009). The effect of repeated questioning on children's accuracy and consistency in eyewitness testimony. *Legal and Criminological Psychology, 14*, 263-278.

Kulhavy, R.W., Yekovich, F.R., & Dyer, J.W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68*, 522-528.

La Rooy, D., Katz, C., Malloy, L.C., & Lamb, M.E. (2010). Do we need to rethink guidance on repeated interviews. *Psychology, Public Policy and Law, 16*, 373-392.

La Rooy, D., Lamb, M.E., & Pipe, M-E. (2009). Repeated interviewing: A critical evaluation of the risks and potential benefits. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (Vol. 327-361). Hoboken, NJ: Wiley.

Lampinen, J.M., Scott, J., Pratt, D., Leding, J.K., & Arnal, J.D. (2007). 'Good, you identified the suspect... But please ignore this feedback': Can warnings eliminate the effects of post-identification feedback? *Applied Cognitive Psychology, 21*, 1037-1056.

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

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1077-1083.

Lindsay, D. S., Allen, B.P., Chan, J. C. K., & Dahl, L.C. (2004). Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another event. *Journal of Memory and Language, 50*, 96-111.

Loftus, E. F. (1979). Reactions to blatantly contradictory information. *Memory & Cognition, 7*, 368-374.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48*, 518-537.

MacLeod, M.D. (2002). Retrieval-induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology, 16*, 135-149.

Marche, T.A. (1999). Memory strength affects reporting of misinformation. *Journal of Experimental Child Psychology, 73*, 45-71.

McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition, 34*, 261-267.

Melnyk, L., & Bruck, M. (2004). Timing moderates the effects of repeated suggestive interviewing on children's eyewitness memory. *Applied Cognitive Psychology, 18*, 613-631.

Memon, A., Holley, A., Wark, L., Bull, R., & Kohnken, G. (1996). Reducing suggestibility in child witness interviews. *Applied Cognitive Psychology, 10*, 503-518.

Memon, A., & Vartoukian, R. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology, 87*, 403-415.

Memon, A., Zaragoza, M., Clifford, B.R., & Kidd, L. (2010). Incoulation or antidote? The effects of cognitive interview timing on false memory for forcibly fabricated events. *Law and Human Behavior, 34*, 105-117.

Odinot, G., Wolters, G., & Lavender, T. (2009). Repeated partial eyewitness questioning causes confidence inflation but not retrieval-induced forgetting. *Applied Cognitive Psychology, 23*, 90-97.

Otani, H., Kato, K., Von Glahn, N.R., Nelson, M.E., Widner, R.L., & Goernert, P.N. (2008). Hypermnesia: A further examination of age differences between young and older adults. *British Journal of Psychology, 99*, 265-278.

Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. Gist testing. *Memory & Cognition, 39*, 155-170.

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

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*, 261-285.

Peterson, C., Parsons, T., & Dean, M. (2004). Providing misleading and reinstatement information a year after it happened: Effects on long-term memory. *Memory, 12*, 1-13.

Register, P.A., & Kihlstrom, J.F. (1988). Hypnosis and interrogative suggestibility. *Personality and Individual Differences, 9*, 549-558.

Roebers, C.M., & Fernandez, O. (2002). The effects of accuracy motivation on children's and adults' event recall, suggestibility, and their answers to unanswerable questions. *Journal of Cognition and Development, 3*, 415-443.

Roediger, H. L., 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.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814.

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

Scrivner, E., & Safer, M.A. (1988). Eyewitness show hypermnesia for details about a violent event. *Journal of Applied Psychology, 73*, 371-377.

Shaw, J.S. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied, 2*, 126-146.

Shaw, J.S., Bjork, R.A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2*(2), 249-253.

Shaw, J.S., & McClure, K.A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior, 20*, 629-653.

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.

Thomas, A.K., Bulevich, J.B., & Loftus, E. F. (2003). Exploring the role of repetition and sensory elaboration in the imagination inflation effect. *Memory & Cognition, 31*, 630-640.

Turtle, J.W., & Yuille, J.C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology, 79*, 260-271.

Wheeler, M. A., Ewers, M., & Buonanno, J.F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*, 571-580.

Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling ballard's (1913) and bartlett's (1932) results. *Psychological Science, 3*, 240-245.

Whitcomb, D., Shapiro, E.R., & Stellwagen, L.D. (1985). When the victim is a child: Issues for judges and prosecutors. Washington, DC: National Institute of Justice.

Wright, D.B. (1993). Misinformation and warnings in eyewitness testimony: A new testing procedure to differentiate explanations. *Memory, 1*, 153-166.

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

Footnotes

[1] Although sex and age data were not collected for participants in the present experiments, we believe that these participants would not differ dramatically from others who we have participated in experiments conducted in our laboratory over the past academic year. Among the 814 subjects tested, the male/female split was 51/49 with a mean age of 19.48 (SD = 2.14, and *range* = 17-39).

Table 1

Mean Probabilities of Correct Recall in the Initial Test Phase, With Standard Deviations Shown in Parentheses

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Experiment 1						
1-test	.51 (.13)					
5-test	.50 (.16)	.53 (.13)	.54 (.12)	.54 (.13)	.54 (.13)	
Experiment 2						
1-test	.61 (.11)					
3-test	.58 (.15)	.61 (.14)	.61 (.13)			
6-test	.54 (.16)	.57 (.17)	.56 (.17)	.55 (.17)	.56 (.18)	.57 (.16)
Experiment 3						
1-test	.51 (.10)					
5-test	.49 (.16)	.53 (.13)	.53 (.13)	.53 (.13)	.54 (.14)	
Experiment 4						
1-test	.59 (.15)					
3-test	.54 (.14)	.56 (.13)	.56 (.14)			
6-test	.56 (.12)	.58 (.12)	.59 (.12)	.59 (.12)	.59 (.13)	.59 (.13)

Table 2

*Mean Probabilities of Recall on The Final Test as a Function of Item Type and Response Type in Experiment 1 (**Between-Subjects, Short Delay**), With Standard Deviations Shown in Parentheses*

	Recall of Original, Correct Item		
	Consistent	Control	Misleading
No-test	.74 (.19)	.47 (.22)	.39 (.20)
1-test	.79 (.19)	.51 (.20)	.28 (.23)
5-test	.86 (.16)	.53 (.17)	.30 (.25)
	Recall of Misinformation		
	Consistent	Control	Misleading
No-test	.01 (.04)	.05 (.09)	.29 (.16)
1-test	.01 (.03)	.06 (.09)	.48 (.30)
5-test	.00 (.02)	.05 (.09)	.55 (.29)
	Other Intrusions		
	Consistent	Control	Misleading
No-test	.20 (.18)	.39 (.19)	.27 (.13)
1-test	.17 (.15)	.36 (.18)	.20 (.16)
5-test	.12 (.14)	.37 (.16)	.14 (.11)
	No Answer		
	Consistent	Control	Misleading
No-test	.05 (.09)	.10 (.10)	.06 (.12)
1-test	.04 (.09)	.07 (.10)	.05 (.09)
5-test	.02 (.05)	.06 (.09)	.02 (.06)

Table 3

*Mean Probabilities of Recall on The Final Test as a Function of Item Type and Response Type in Experiment 2 (**Within-Subjects, Short Delay**), With Standard Deviations Shown in Parentheses*

	Consistent		Control		Misleading	
	Nontested	Tested	Nontested	Tested	Nontested	Tested
Recall of Original, Correct Item						
1-test	.71 (.16)	.83 (.14)	.56 (.23)	.60 (.18)	.38 (.25)	.44 (.28)
3-test	.77 (.16)	.84 (.16)	.55 (.19)	.59 (.22)	.38 (.23)	.32 (.24)
6-test	.68 (.19)	.81 (.17)	.54 (.21)	.59 (.22)	.40 (.27)	.31 (.28)
Recall of Misinformation						
1-test	.03 (.07)	.02 (.06)	.01 (.04)	.04 (.07)	.35 (.24)	.40 (.29)
3-test	.01 (.05)	.01 (.05)	.03 (.09)	.04 (.08)	.39 (.26)	.48 (.29)
6-test	.04 (.08)	.01 (.05)	.06 (.10)	.05 (.08)	.29 (.21)	.55 (.34)
Other Intrusions						
1-test	.19 (.15)	.12 (.11)	.34 (.19)	.29 (.18)	.22 (.16)	.12 (.12)
3-test	.18 (.15)	.14 (.14)	.36 (.20)	.33 (.20)	.19 (.16)	.19 (.15)
6-test	.21 (.18)	.15 (.11)	.33 (.18)	.31 (.19)	.25 (.18)	.12 (.15)
No Answer						
1-test	.06 (.09)	.02 (.06)	.09 (.15)	.07 (.12)	.06 (.14)	.03 (.07)
3-test	.03 (.08)	.01 (.04)	.07 (.11)	.04 (.11)	.05 (.09)	.01 (.04)
6-test	.07 (.11)	.03 (.12)	.07 (.13)	.06 (.13)	.07 (.12)	.03 (.09)

Table 4

*Mean Probabilities of Recall on The Final Test as a Function of Item Type and Response Type in Experiment 3 (**Between-Subjects, Long Delay**), With Standard Deviations Shown in Parentheses*

	Consistent	Control	Misleading
Recall of Original, Correct Item			
No-test	.63 (.21)	.19 (.13)	.09 (.09)
1-test	.80 (.11)	.46 (.17)	.24 (.21)
5-test	.83 (.17)	.55 (.18)	.25 (.24)
Recall of Misinformation			
No-test	.03 (.06)	.05 (.08)	.44 (.16)
1-test	.00 (.02)	.04 (.07)	.51 (.22)
5-test	.01 (.04)	.05 (.07)	.59 (.24)
Other Intrusions			
No-test	.24 (.13)	.53 (.21)	.31 (.19)
1-test	.15 (.11)	.40 (.19)	.20 (.13)
5-test	.14 (.15)	.34 (.16)	.16 (.13)
No Answer			
No-test	.11 (.11)	.24 (.19)	.16 (.14)
1-test	.05 (.08)	.11 (.11)	.06 (.09)
5-test	.02 (.05)	.06 (.08)	.01 (.04)

Table 5

*Mean Probabilities of Recall on The Final Test as a Function of Item Type and Response Type in Experiment 4 (**Within-Subjects, Long Delay**), With Standard Deviations Shown in Parentheses*

	Consistent		Control		Misleading	
	Nontested	Tested	Nontested	Tested	Nontested	Tested
Recall of Original, Correct Item						
1-test	.64 (.21)	.72 (.27)	.32 (.23)	.56 (.18)	.25 (.23)	.44 (.28)
3-test	.58 (.22)	.83 (.16)	.25 (.21)	.58 (.23)	.20 (.18)	.38 (.29)
6-test	.62 (.26)	.81 (.21)	.30 (.16)	.51 (.23)	.17 (.17)	.32 (.27)
Recall of Misinformation						
1-test	.04 (.08)	.04 (.09)	.08 (.18)	.07 (.13)	.41 (.27)	.37 (.26)
3-test	.05 (.09)	.01 (.06)	.11 (.13)	.05 (.09)	.44 (.27)	.43 (.27)
6-test	.05 (.08)	.03 (.07)	.08 (.11)	.05 (.10)	.46 (.19)	.53 (.31)
Other Intrusions						
1-test	.26 (.20)	.20 (.22)	.46 (.22)	.32 (.18)	.25 (.20)	.16 (.19)
3-test	.28 (.19)	.15(.14)	.45 (.20)	.28 (.20)	.25 (.17)	.18 (.17)
6-test	.25 (.20)	.13 (.19)	.47 (.23)	.35 (.21)	.29 (.20)	.14 (.15)
No Answer						
1-test	.06 (.10)	.04 (.11)	.14 (.20)	.05 (.10)	.10 (.15)	.04 (.07)
3-test	.10 (.15)	.01 (.04)	.19 (.20)	.09 (.12)	.12 (.13)	.02 (.05)
6-test	.08 (.12)	.03 (.07)	.15 (.18)	.08 (.13)	.09 (.11)	.02 (.06)

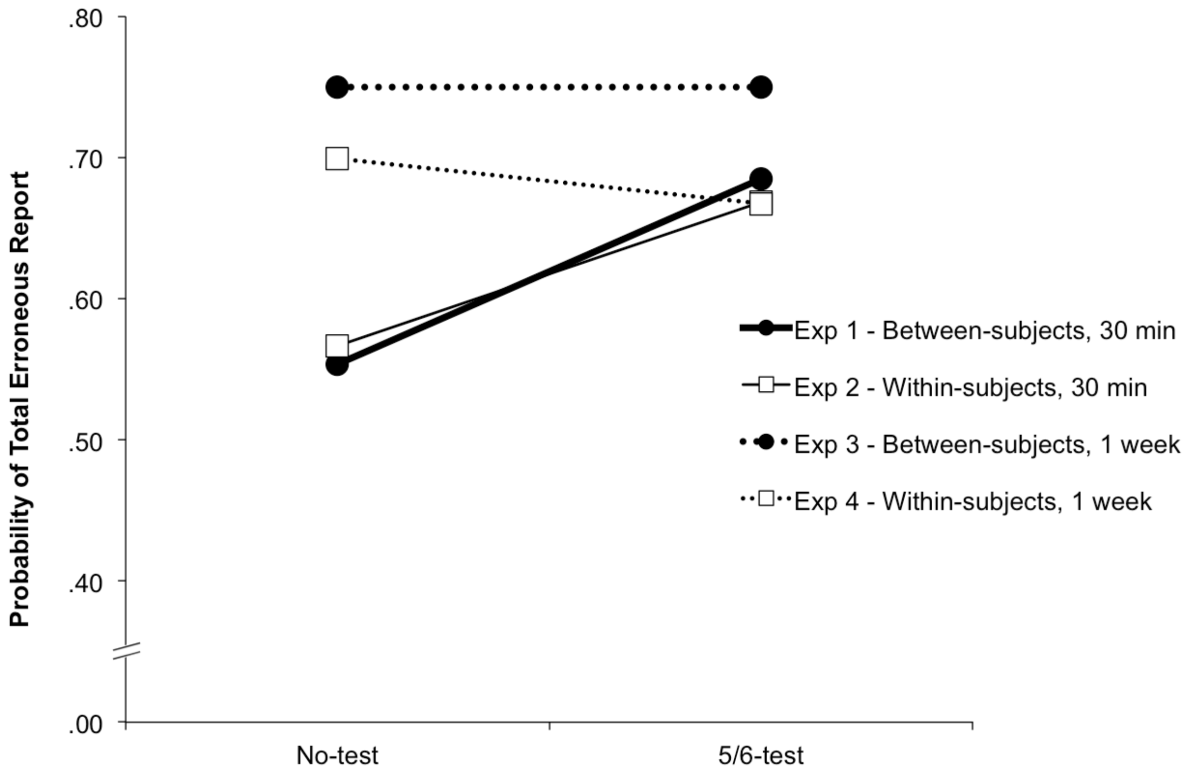

Figure 1. Mean probabilities of total commission errors (i.e., Misinformation + Other Intrusions) during the final recall test of the four experiments as a function of number of initial tests. The no-test error rates for the within-subjects experiments (Exp 2 and 4) were based on the averaged recall probabilities of the nontested items in the 1-test, 3-test, and 6-test conditions.

Appendix

The 36 Questions, Along With Their Correct Answer and Misinformation, Used in the Present Experiments. Items With an Asterisk Were Used in All Four Experiments; Items Without an Asterisk Were Used in Only Experiments 1 and 3.

1. What is David Palmer's shirt color? [Correct Answer: Blue, Misinformation: Grey]
2. What is Sheri Palmer working on at the beginning of the episode? * [Correct Answer: Thank you notes, Misinformation: Campaign staff appointments]
3. How many floors does Jack Bauer's house have? [Correct Answer: One, Misinformation: Two]
4. What are Jack and Kim doing at the beginning of the episode? [Correct Answer: Playing chess, Misinformation: Playing Backgammon]
5. At the beginning of the episode, what color is Jack Bauer's shirt? * [Correct Answer: Blue, Misinformation: White]
6. When you first see her, what is Teri Bauer's shirt color?* [Correct Answer: Green, Misinformation: Blue]
7. What item does Jack take out of the refrigerator?* [Correct Answer: Jello Cup, Misinformation: Soda]
8. What band poster is on Kim's bathroom door?* [Correct Answer: Coldplay, Misinformation: Radiohead]
9. How long does Jack expect to be gone when he is called from home? * [Correct Answer: 1 hour, Misinformation: 2 hours]
10. Jack speaks to Kim's ex-boyfriend on the phone. What is his name? * [Correct Answer: Vincent, Misinformation: Xander]

11. What type of vehicle does Jack drive? * [Correct Answer: SUV, Misinformation: Pick-up truck]
12. Janet and Kim are meeting two guys. What year are they in school? [Correct Answer: Sophomores, Misinformation: Freshmen]
13. What item is on the dashboard of Janet's car? * [Correct Answer: Hawaiian Lei, Misinformation: Fuzzy dice]
14. What color is Nina's shirt? * [Correct Answer: Black, Misinformation: Green]
15. What is the color of Dan & Rick's van? * [Correct Answer: Purple, Misinformation: Black]
16. How many corrupt agents did Jack investigate at CTU? [Correct Answer: Three, Misinformation: Five]
17. What time is Martin's meeting with Palmer? * [Correct Answer: 7:00 AM, Misinformation: 9:00 AM]
18. What is the name of the woman sitting next to Martin? [Correct Answer: Mandy, Misinformation: Marcy]
19. What college does Rick attend? * [Correct Answer: San Diego State, Misinformation: USC]
20. How long does Kim say it has been since her father died? * [Correct Answer: 6 months, Misinformation: 3 months]
21. Midway through the episode, Jack changes his shirt. What is the new shirt color? [Correct Answer: Grey, Misinformation: Black]
22. What does Mason call when Jack asks for additional approval? * [Correct Answer: Time and Date service, Misinformation: Operator]
23. What does Jack take to cover the tranq gun? [Correct Answer: Binder, Misinformation: Briefcase]

24. Jack talks to Mason about a drug dealer. What type of drug was being trafficked? * [Correct Answer: Heroin, Misinformation: Cocaine]
25. How much money was missing from the drug deal? [Correct Answer: $200,000, Misinformation: $300,000]
26. What is Kim's password for her email? [Correct Answer: LIFESUCKS, Misinformation: SCREWLIFE]
27. What city has Martin recently photographed? * [Correct Answer: Munich, Misinformation: Frankfurt]
28. What part of the city does Teri tell Jack that Kim has gone? * [Correct Answer: Valley, Misinformation: Downtown]
29. How many messages does Kim have on her cell phone? [Correct Answer: Five, Misinformation: Four]
30. Where is George Mason's illicit bank account? * [Correct Answer: Aruba, Misinformation: Costa Rica]
31. What does the terrorist use on the flight attendant? * [Correct Answer: Hypodermic syringe, Misinformation: Chloroform rag]
32. What is the explosive with the timer hidden in? * [Correct Answer: Fire extinguisher, Misinformation: Drink cart]
33. What color is the car Teri is in at the end of the episode? [Correct Answer: Silver, Misinformation: Blue]
34. What type of plane does Tony report has exploded? * [Correct Answer: 747, Misinformation: DC-10]

35. Which desert does Tony report that the plane exploded over? * [Correct Answer: Mojave Desert, Misinformation: Baja Desert]
36. What street does Kim say she lives on when she is in Dan's van? * [Correct Answer: 10^{th} Street, Misinformation: 25^{th} Street]

Supplementary Materials

Table A.

*Mean Probabilities of Recall on The Modified-Modified Free Recall (MMFR) Test as a Function of Item Type and Response Type in Experiment 4 (**Within-Subjects, Long Delay**), With Standard Deviations Shown in Parentheses*

	Consistent		Control		Misleading	
Recall of Original, Correct Item						
	Nontested	Tested	Nontested	Tested	Nontested	Tested
1-test	.71 (.21)	.82 (.18)	.33 (.21)	.65 (.18)	.31 (.23)	.60 (.22)
3-test	.69 (.22)	.83 (.20)	.37 (.25)	.60 (.23)	.29 (.24)	.51 (.29)
6-test	.65 (.27)	.82 (.23)	.37 (.19)	.57 (.23)	.29 (.33)	.51 (.26)
Recall of Misinformation						
	Consistent		Control		Misleading	
	Nontested	Tested	Nontested	Tested	Nontested	Tested
1-test	.11 (.14)	.07 (.11)	.12 (.17)	.10 (.12)	.59 (.22)	.63 (.24)
3-test	.08 (.12)	.05 (.11)	.15 (.17)	.06 (.11)	.53 (.26)	.62 (.23)
6-test	.11 (.19)	.09 (.17)	.13 (.14)	.07 (.10)	.50 (.25)	.60 (.28)
Other Intrusions						
	Consistent		Control		Misleading	
	Nontested	Tested	Nontested	Tested	Nontested	Tested
1-test	.38 (.23)	.29 (.18)	.61 (.19)	.43 (.20)	.36 (.23)	.28 (.22)
3-test	.38 (.21)	.26 (.23)	.59 (.28)	.41 (.24)	.37 (.25)	.29 (.21)
6-test	.37 (.25)	.22 (.20)	.59 (.22)	.45 (.31)	.40 (.24)	.24 (.19)
No Answer						
	Consistent		Control		Misleading	
	Nontested	Tested	Nontested	Tested	Nontested	Tested
1-test	.02 (.06)	.03 (.07)	.10 (.18)	.00 (.00)	.06 (.11)	.02 (.05)
3-test	.04 (.09)	.06 (.29)	.12 (.30)	.08 (.32)	.09 (.29)	.01 (.03)
6-test	.02 (.05)	.02 (.06)	.07 (.13)	.05 (.11)	.03 (.08)	.03 (.11)