Are witnesses able to avoid highly accessible misinformation? Examining the efficacy of different warnings for high and low accessibility postevent misinformation

John B. Bulevich[1] · Leamarie T. Gordon[2] · Gregory I. Hughes[3] · Ayanna K. Thomas[3]

Abstract

The reliability of eyewitness memory continues to be an area of concern, particularly in situations that involve conflicting sources of information (e.g., the misinformation effect; Loftus et al., 1978, *Journal of Experimental Psychology: Human Learning and Memory, 4*[1], 19–31). To mitigate the negative effects of misinformation, researchers have examined the efficacy of warnings that highlight the unreliability of postevent information. However, warnings have proven less effective for highly accessible misinformation (Eakin et al., 2003, *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29[5], 813–825). In the present study, we examined the effects of different types of warnings for low accessibility misinformation in a standard single test misinformation paradigm, and highly accessible misinformation in a repeated testing misinformation paradigm (Chan et al., 2009, *Psychological Science*, 20[1], 66–73). We modeled these warnings after Eakin et al. (2003) to include both general warnings and specific question-by-question warnings. We found that warnings were effective in both types of misinformation paradigms. Additionally, memory accuracy in situations where participants were exposed to misleading information was improved when specific and general warnings were combined. We argue that both retrieval blocking of low accessibility items and enhanced contextual discrimination account for these findings.

Keywords Memory · Eyewitness testimony · Metamemory

Eyewitness memory has been shown to be unreliable, especially in situations where postevent information is introduced (Frenda et al., 2011). Eyewitness memory is frequently investigated using the misinformation paradigm, in which participants recollect details from a complex event after exposure to misleading postevent information. Findings from this methodology have consistently demonstrated that when postevent information contains details that are inconsistent with the originally witnessed event, memory for the original event is impaired (for review, see Frenda et al., 2011). One theory to explain why individuals might report misinformation as their own memories assumes that when individuals are exposed to postevent information, accessibility of that information increases while access to original

event details is blocked (e.g., Belli et al., 1992; Eakin et al., 2003).

Although research has demonstrated that postevent information negatively impacts reporting of original event details, it has also demonstrated that postevent information may not necessarily eliminate access to original event details altogether (e.g., Lindsay & Johnson, 1989; McCloskey & Zaragoza, 1985). The accessibility of postevent information has been shown to be a significant determinant in whether memory for the original event is impaired (see Loftus et al., 1978, for an early example). Given the likelihood that original event details may remain accessible, researchers have examined whether warnings about the quality of postevent information may promote increased scrutiny at final test, resulting in improved access to and reporting of original event details.

The primary focus of the present research was to examine whether the efficacy of warnings interacted with the accessibility of the postevent misleading information. Specifically, we examined how warnings affected misinformation accuracy across two experiments.

✉ John B. Bulevich
  john.bulevich@stockton.edu

[1] Stockton University, Galloway, NJ, USA

[2] Assumption University, Worcester, MA, USA

[3] Tufts University, Medford, MA, USA

Research on warnings and the eyewitness misinformation effect have been explored for several decades (see Greene et al., 1982, for an early example, and Echterhoff et al., 2005, for a more recent one). Warnings can be delivered prior to misinformation (Karanian et al., 2020) or afterwards (see Blank & Launay, 2014, for a review and meta-analysis). Warnings are typically presented through experimental instructions that encourage caution during retrieval (Thomas et al., 2010). For our purposes, we have focused on warning effects in the context of the traditional three-phase eyewitness misinformation paradigm (Loftus et al., 1978), where participants witness a mock crime, and are exposed to misleading postevent information about that crime.

For the present research, we compared two different types of warnings given after the presentation of misleading postevent information and designed to highlight conflicts between the original event and postevent narrative. The first warning type was similar to the warning presented to participants in Greene et al. (1982), used by Lindsay (1990), where participants were told that the narrative may include inaccurate information and that they should try to answer questions based on their memory for the original event. The second was modeled after the modified opposition test (MOT)[1] used by Eakin et al. (2003). In this context, we presented participants with a hint about the answer to a test question on an item-by-item basis (What did the man use to open the window? [Hint: it was not a wrench]). Similar research examining item-by-item warnings (Higham et al., 2017), which was designed to explore situations where participants detected discrepancies between an original event and misinformation, specified which test questions involved exposure to misleading information but stopped short of actually providing the misleading detail. In a standard misinformation paradigm, Eakin et al. found that both the general post-misinformation warning and the modified opposition warnings were effective, but only in situations where the misleading items were defined as low accessibility. They defined misleading items as low accessibility if that information was presented only one time in the postevent narrative. In contrast, they also had a condition where the misleading items were presented in the narrative *and* occurred on a recognition test that preceded the final cued recall test. Critically, this two-alternative forced-choice (2AFC) recognition test presented the misleading item and a novel foil item. These choices were used to promote misinformation selection. Eakin et al. termed this

the high accessibility condition, and they found that warnings were ineffective under these circumstances.

In a parallel line of research Chan et al. (2009) gave participants a memory test on the original event immediately following witnessing the original event. They hypothesized that retrieval in this context would increase accessibility of original event details and reduce susceptibility to postevent misinformation. However, counter to their original predictions, Chan et al. found that testing after witnessing an event increased susceptibility to misleading postevent information, a finding later termed *retrieval enhanced suggestibility* (RES; Thomas et al., 2010). Since this initial work, the RES effect has been replicated in more than a dozen studies (e.g., Brackmann et al., 2016; Butler & Loftus, 2018; Chan & Langley, 2011; Gordon et al., 2015; Gordon & Thomas, 2014, 2017; LaPaglia et al., 2014; Manley & Chan, 2019; Rindal et al., 2016; Thomas et al., 2010; Thomas et al., 2017; Wilford et al., 2014).

While initial testing has been shown to increase misinformation susceptibility, warnings provided before a final test have been shown to mitigate RES. In an early investigation of RES, Thomas et al. (2010) gave participants an initial test of the original event and then exposed them to misinformation in a narrative. Half of the participants were given a general post-misinformation warning that the narrative may not be reliable. As a reminder, this type of warning is similar to the warning from Greene et al. (1982), where participants were told that some of the information in the narrative may be incorrect. The authors found that under these conditions, the warning reduced susceptibility to misleading information and eliminated the RES effect. The authors concluded that warned participants analyzed their retrieval output more closely than those not warned. This conclusion was supported by data demonstrating that participants who were warned made recognition decisions on a 4AFC test related to misleading details slower than those who were not warned.

In addition, for unwarned participants in the RES group who were given the recognition test that included the misleading detail and the original event detail, they selected the misleading information faster than participants who had not taken an initial test of the witnessed event. Thomas et al. argued that these data suggested that the misleading information was highly accessible.

Thomas et al. (2010) further argued that the RES group had robust memory representations for both the original event and the postevent information. Because both memories were strong, participants were better able to use contextual cues to discriminate between sources of information. More contextual cues may also have facilitated more effective retrieval monitoring (cf. Gallo, 2004). That is, the initial test present in the RES method may result in a memory representation for the original event with richer, more accessible contextual detail compared with when the original event is

---

[1] The MOT was originally conceived as a complement to the opposition warning created by Lindsay (1990) where participants were instructed to ignore *all* information presented in the narrative. In the present research, the narrative contains both correct and incorrect information.

not tested (cf. Thomas et al., 2017). Concurrently, the initial test may potentiate learning of the postevent information presented in the narrative (e.g., Gordon & Thomas, 2017). Some research using the RES methodology has found evidence that taking a test of the original event changes how participants process the postevent narrative. For example, Gordon and Thomas (2014) found that participants spent more time reading misleading details if they had taken an initial test of the original event. Gordon and Thomas interpreted these findings within the context of a forward effect of testing (cf. Pastötter & Bäuml, 2014). That is, an initial test directly affects processing of related subsequently presented material. Regardless of the accuracy, postevent information may be better learned if a test precedes it.

Consistent with these ideas Gordon and Thomas (2014) found that when participants were asked to retrieve as many details as they could remember on a final memory test, participants in the RES group (initial and final test) were more likely to retrieve original details *and* postevent misleading details from the narrative as compared with participants in the single test group. Similarly, when participants were required to retrieve only narrative details, participants who took an initial test retrieved more narrative details than those who did not (Gordon & Thomas, 2017). Thus, when paired with a warning that encouraged additional scrutiny of retrieved details, participants may be better able to discriminate between the two sources of information. We argued that the RES paradigm leads to a more contextually rich memory for the original event. However, RES also makes postevent misinformation extremely accessible. Thus, it is only when people are given a prompt or warning to more closely examine their responses will they scrutinize their memory for the original event.

The results of Thomas et al. (2010) indicate that warnings might be an effective means to encourage effective retrieval monitoring, even when postevent information is highly accessible, and seem to contrast with the findings of Eakin et al. (2003). Within the context of RES, increasing accessibility of original *and* postevent details increased misinformation susceptibility; however, that susceptibility could be ameliorated using a general warning. Alternatively, Eakin et al. (2003) found the same type of warning was ineffective in reducing misinformation effects when accessibility to post6event misleading details was increased through retrieval of the details following narrative encoding.

In the context of the misinformation paradigm, we conceptualized warnings as an exogenous prompt that encourages the exercise of metacognitive control and the inspection of the source of retrieved information to improve accuracy. While not the focus of the present research, we also measured retrospective confidence ratings to explore the relationship between warnings and confidence. Research has consistently demonstrated overconfidence

after exposure to misleading postevent information (Loftus et al., 1989; Weingardt et al., 1994). However, the relationship between confidence and accuracy in misinformation experiments has been positively affected by general warnings (e.g., Thomas et al., 2010) and retrieval effort (Bulevich & Thomas, 2012). However, Higham et al. (2017) found in their related research that only warnings that specified which test items corresponded to exposure to misleading information produced a more accurate relationship between memory performance and confidence. The present set of experiments further explored the role of confidence for high and low accessible misinformation, and when participants were exposed to a general warning given prior to the final test or a final test that instantiated item-level warnings.

In addition, the present study served as a conceptual replication of Eakin et al. (2003) using a standard misinformation design, and to examine the effectiveness of different types of warning within the context of the RES paradigm, where original and postevent details are rendered highly accessible. In Experiment 1, we used a standard misinformation paradigm that did not include a retrieval practice phase to examine these variables in a situation where postevent narrative details are less accessible. In Experiment 2, we increased the accessibility of postevent misinformation by using the RES methodology and incorporating initial testing of the original event prior to presentation of the narrative. Under these conditions, warnings should be less effective in reducing misinformation susceptibility. In both experiments, warnings were general, specific to misleading items, or a combination of both warnings.

For Experiment 1, we predict that, in a standard misinformation paradigm, the results should replicate the findings for low accessibility items in Eakin et al. (2003). Specifically, we expect that both kinds of warnings would reduce susceptibility to misleading postevent information. For Experiment 2, where we examined the impact of these two kinds of warning on misinformation susceptibility in the RES paradigm, the predictions are less clear. There are two primary possibilities. The first is that misleading information in the RES paradigm will behave similarly to the highly accessible information in Eakin et al., and therefore, warnings will be ineffective. Since the initial test has been shown to increase the accessibility of postevent information (e.g., Gordon & Thomas, 2014, 2017), neither the general nor the specific warning should reduce misinformation susceptibility. However, as described earlier, initial testing in RES has also been shown to increase the accessibility of the original event (Gordon & Thomas, 2017). In the context of the present research, warnings may encourage individuals to engage in retrieval monitoring processes that capitalize on the increased accessibility of both original event and postevent information.
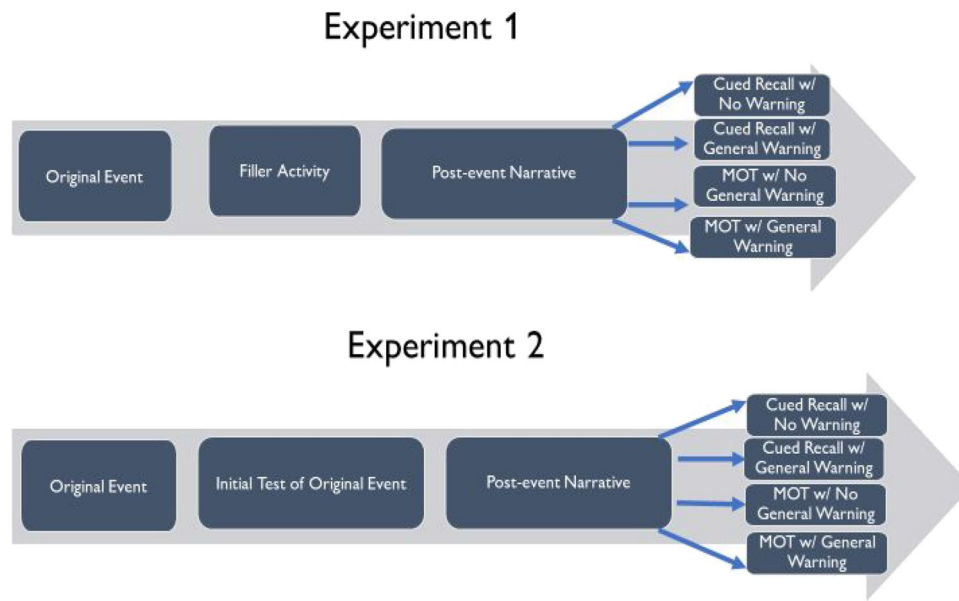
Experiment 1

Experiment 2

Fig. 1 Overview of the procedures for Experiment 1 and Experiment 2

Experiment 1

In Experiment 1, participants were exposed to a witnessed event and then to a postevent narrative. The narrative included misleading details (inconsistent with the event) and critical neutral details. We compared participants who received a general warning immediately before the final test with those who did not receive a warning. Further, participants either took a final standard cued recall test or a cued recall that included item level warnings (MOT; see Fig. 1 for an overview of the procedure). We predicted that both the general warning and item-level warnings would reduce misinformation susceptibility, as measured by the comparison in accuracy between neutral and misleading trials.

Method

Participants

We conducted sensitivity power analyses to determine the minimum detectable effect sizes that our sample could detect (Cohen, 1988; for a discussion, see Giner-Sorolla et al., 2019). The analysis procedure involves solving for an effect size given known values of α, power $(1 - \beta)$, and the degrees of freedom (df) of a statistical test. In our three-way mixed design, a minimum detectable effect size can be calculated for each of the tests of the main effects and interactions of the within-subjects and between-subjects variables.

Our effect-size sensitivity analysis procedure involved two steps. First, we used the generic F-test module[2] of the G*Power software (Faul et al., 2007) to calculate a noncentrality parameter (λ), which requires choosing a desired level of α (.05), power, and the degrees of freedom of the numerator and denominator of the statistical test of interest. Second, we used the noncentrality parameter to solve for the minimum detectable effect size using the formulas outlined in Faul et al. (2007; Table 3),[3] which incorporates the correlation among repeated measures (r = .71). The results of these effect-size sensitivity analyses are shown in Table 1 for two common levels of desired statistical power (80% and 90%).

Here, we focus on the two statistical tests that were most pertinent to our hypotheses—the pair of two-way interactions of the within-subjects variable (item type) and each of the between-subjects variables (final test, warning group). At a power level of 90%, the minimum detectable effect size (η_p^2) of these tests is .0097, which is slightly less than the standard cutoff for what is considered a small effect (.01; Cohen,

---

[2] G*Power does not readily accommodate three-way mixed designs. With our design, power analyses for the lower-order tests (e.g., a 2 × 2 interaction) are problematic because G*Power uses the wrong degrees of freedom, leading to incorrect results. The generic F-test option allows the appropriate flexibility to specify the correct degrees of freedom. Users of G*Power should consult Lakens (2013) on how to avoid additional pitfalls of the software.

[3] The formulas in Faul et al. (2007) use Cohen's f as the effect size. Per Cohen (1988), converting f to η_p^2 can be accomplished through the formula, $\eta_p^2 = \frac{f^2}{1+f^2}$

Table 1 Experiment 1: The results of the sensitivity power analyses and the observed effect sizes

Statistical test	Power $(1 - \beta)$	Minimum detectable effect size (η_p^2)	Observed effect size (η_p^2)
Main effect: Item type	.80	.007	.060
	.90	.010	
Main effect: Warning group	.80	.042	.032
	.90	.056	
Main effect: Final test	.80	.042	.118
	.90	.056	
Interaction: Item Type × Warning Group	.80	.007	.030
	.90	.010	
Interaction: Item Type × Final Test	.80	.007	.036
	.90	.010	
Interaction: Warning Group × Final Test	.80	.007	.001
	.90	.010	
Interaction: Item Type × Warning Group × Final Test	.80	.007	.002
	.90	.010	

Table 2 Experiment 2: The results of the sensitivity power analyses and the observed effect sizes

Statistical test	Power $(1 - \beta)$	Minimum detectable effect size (η_p^2)	Observed effect size (η_p^2)
Main effect: Item type	.80	.008	.199
	.90	.010	
Main effect: Warning group	.80	.034	.038
	.90	.045	
Main effect: Final test	.80	.034	.079
	.90	.045	
Interaction: Item Type × Warning Group	.80	.008	.054
	.90	.010	
Interaction: Item Type × Final Test	.80	.008	.099
	.90	.010	
Interaction: Warning Group × Final Test	.80	.034	.000
	.90	.045	
Interaction: Item Type × Warning Group × Final Test	.80	.034	.018
	.90	.045	

1988). In this experiment, we observed effect sizes that were several times larger than the minimum detectable effect size, both of which can be considered small-to-medium effects (η_p^2 = .04 [Item Type × Final Test] and η_p^2 = .03 [Item Type × Warning Group]). Small-to-medium effect sizes are typical in the broader literature on warnings and the misinformation effect (for a meta-analysis, see Blank & Launay, 2014; Table 2).

A total of 155 undergraduate students volunteered to participate in this study for course credit (M_{age} = 20.01, SD_{age} = 2.88). Students were recruited from Stockton University (N =105; M_{age} = 19.61, SD_{age} = 2.26) and Tufts University (N =50; M_{age} = 20.20, SD_{age} = 3.11). Nine participants were excluded from the analysis for failing to comply with stated instructions or because of data corruption. Participants were randomly assigned to one of four between-subjects groups: cued recall no warning (n = 42; all from Stockton University), cued recall and a warning (n = 34; 19 from Stockton University), MOT no warning (n = 43; 25 from Stockton University), or MOT with a warning (n = 36; 19 from Stockton University).

Design

A 2 (final test: cued recall, MOT) × 2 (warning group: no warning, warning) × 2 (item type: neutral, misleading) mixed design was used in Experiment 1. Final test and warning group was manipulated between participants, and item type was manipulated within participants.

Materials and procedure

The procedure consisted of watching a video of a crime, a filler task, postevent narrative processing, and warning paired with a final memory test. Participants were tested individually or in groups of one to four on computers programmed with E-Prime software (Version 2.1; Schneider et al., 2002). All sessions began with informed consent and a basic demographic questionnaire.

Participants watched a 24-min portion of a black-and-white, silent film titled *Rififi* (Bezard et al., 1955). The video depicted four men committing a breaking-and-entering burglary. No participant reported having seen the video before. Twenty-four details from the video served as items on subsequent memory tests (materials adapted from Gordon & Thomas, 2017).

After an 8-min filler task in which participants played the computerized game Tetris (Nintendo of America, Redmond, WA, USA), participants in all groups listened to a 10-min, 22-s narrative synopsis of the video. Participants were instructed to read a verbatim transcript of the narrative as they listened. The narrative consisted of 116 sentences, 24 of which referenced each of the 24 details from the memory tests (e.g., The last assailant took a *ring* out of a drawer). These 24 sentences were split into two sets of 12. In one set, the critical detail in each sentence was inconsistent with the video (e.g., the last assailant took a *necklace* out of a drawer), which we termed *misleading items*. In the other set, the critical detail in each sentence provided nonspecific information compared with the video (e.g., the last assailant took *jewelry* out of a drawer), which we termed *neutral items*. These two sets were counterbalanced across participants. No information in the other sentences was assessed during the memory tests.

Immediately following the narrative, participants in two warning groups (those who received the warning alone or in combination with the MOT) received a warning about the postevent narrative. Participants in these two groups were told that some information in the narrative may have been inaccurate, and therefore they should answer questions on the subsequent memory test only from what they remembered seeing in the video. Participants read this warning immediately before taking the final memory test. Participants in the MOT and no warning group were simply told they would be taking a memory test regarding the video.

During the next part of the procedure, all participants were informed that they would take a memory test concerning details in the video. Participants who took the cued-recall test were asked to answer 24 cued-recall questions concerning specific details from the video (e.g., The last assailant took a _____ out of a drawer). Each question remained on the screen for 15 s, during which participants could type an answer. After the 15 s elapsed, the screen advanced to the next question. Participants could not advance to the next question before the 15 s time was completed, nor could they revisit prior questions. However, participants could choose not to answer a given question. Unanswered questions were scored as incorrect. After answering each question, participants were asked to provide confidence in their answer on a 0 (*complete guess*) to 100% (*completely confident*) scale, and there was no time limit to respond. Questions were presented in the same order as the corresponding details that appeared in the video.

Participants who were given the MOT were given the same cued-recall test and also given explicit instructions to ignore misleading details that had been presented in the narrative. These instructions were given on a trial-by-trial basis in which a question would be presented (e.g., What did the assailant remove from the drawer?) and directly below, participants would be warned that the answer was not the incorrect detail presented in the narrative (e.g., the answer is not *necklace*. Do not provide *necklace* as your answer). We provided the specific warning both for misleading and neutral items in the narrative. Therefore, participants were exposed to all misleading details in the context of this MOT.

To summarize, all groups of participants watched the video (original event), engaged in a short filler task (Tetris), and were presented with a synopsis of the video that included misleading details. Following the narrative, some participants received a general warning about the narrative, whereas some did not (warning vs. no warning). Additionally, some participants received a standard cued recall test and others received the MOT version of the final test.

Results

Final test accuracy

We used accuracy as our primary measure of memory performance as related to misinformation impairment. While some research (including some by the present authors; Thomas et al., 2010) have utilized misinformation production (errors of commission involving misleading information), we used accuracy because the MOT provided the misleading information as part of the warning. Thus, consistent with a large body of prior research, we defined misinformation susceptibility as the impairment in retrieval of correct information after the presentation
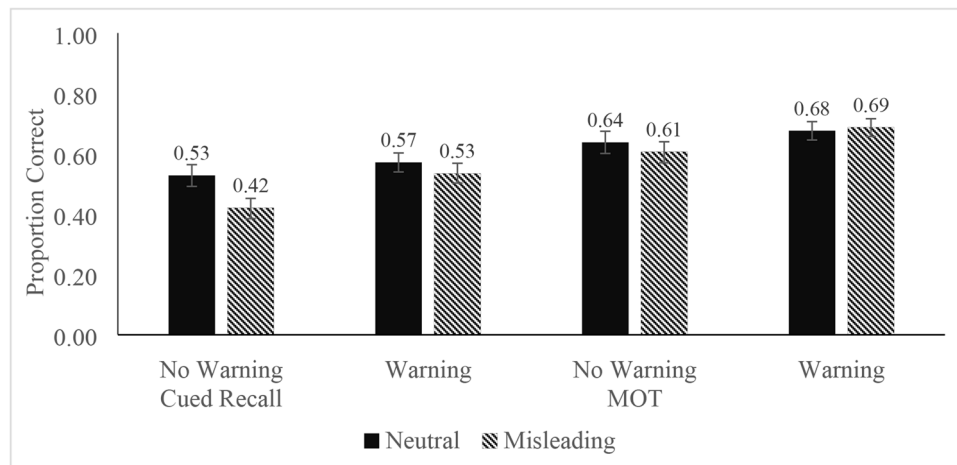
Fig. 2 Accuracy on the final test as a function of item type, warning group, and type of final test for Experiment 1. (*M* and *SE* plotted)

of misleading details (see Belli et al., 1994; McCloskey & Zaragoza, 1985, for early examples). Further, by defining misinformation susceptibility in this way, the possible alternative explanation of experimental demand as an explanation for the misinformation effect is reduced.

Accuracy was calculated by taking the proportion correct on neutral trials and misleading trials for each participant. The proportion correct within each class of items for each participant was then analyzed in a 2 (final test: cued recall, MOT) × 2 (warning group: no warning, warning) × 2 (item type: neutral, misleading). Final test and warning group were between-subject variables, and item type was a within-subjects variable. Consistent with the large body of research demonstrating the misinformation effect, we found a main effect of item type, $F(1, 151) = 9.67$, $p = .002$, $\eta_p^2 = .06$. Participants were more accurate on neutral trials ($M = .60$, $SE = .02$) as compared with misleading trials ($M = .56$, $SE = .02$). We also found a main effect of warning group, $F(1, 151) = 5.05$, $p = .026$, $\eta_p^2 = .03$. Participants were more accurate after receiving a warning ($M = .62$, $SE = .02$) as compared with no warning ($M = .55$, $SE = .02$). Finally, a main effect of final test was found, $F(1, 151) = 20.28$, $p < .001$, $\eta_p^2 = .12$. Participants were more accurate after they had taken the MOT ($M = .65$, $SE = .02$) as compared with a standard cued recall test ($M = .51$, $SE = .02$). Although final test did not interact with warning group, $F < 1$, item type interacted with both variables, Item Type × Final Test, $F(1, 151) = 5.61$, $p = .019$, $\eta_p^2 = .04$; Item Type × Warning Group, $F(1, 151) = 4.63$, $p = .033$, $\eta_p^2 = .03$. As Fig. 2 illustrates, the type of final test and warning has greater effects on misleading trials, $t(153) = 2.82$, $p = .005$, $d = .46$, as compared with neutral trials, $t(153) = 1.17$, $p = .242$.

Withholding

We also examined whether the pattern of withholding responses on the final test would be impacted by type of final test or warning. We conducted a 2 (final test: cued recall, MOT) × 2 (warning group: no warning, Warning) × 2 (item type: neutral, misleading) mixed-design analysis of variance (ANOVA) on average proportion of withholding responses on the final test. We found a main effect of warning group, $F(1, 151) = 7.14$, $p = .008$, $\eta_p^2 = .04$. Participants were less likely to withhold responses after receiving a warning ($M = .11$, $SE = .01$) as compared with no warning ($M = .16$, $SE = .02$). A main effect of final test was also found, $F(1, 151) = 5.59$, $p = .019$, $\eta_p^2 = .04$. Withholding responses was more likely if participants took the MOT ($M = .16$, $SE = .02$) as compared with the cued recall test ($M = .11$, $SE = .01$). Finally, we found an interaction between item type and final test, $F(1, 151) = 7.14$, $p = .008$, $\eta_p^2 = .05$. Participants were more likely to withhold a response associated with a neutral trial ($M = .13$, $SE = .02$) as compared with a misleading trial ($M = .09$, $SE = .02$) if they had taken a cued recall test. However, if participants had taken the MOT, they were more like to withhold responses associated with misleading trials ($M = .17$, $SE = .03$) as compared with neutral trials ($M = .15$, $SE = .03$). These results are not surprising given that each trial of the MOT included information that participants should exclude from their responses.

Confidence

Warning also had an impact on average confidence ratings. We conducted a 2 (final test: cued recall, mot) × 2 (warning group: no warning, warning) × 2 (item type: neutral, misleading) mixed design ANOVA on average confidence

associated with answers on the final test. We found an interaction between item type and final test, $F(1, 150) = 4.21$, $p = .042$, $\eta_p^2 = .03$. In addition, we found an interaction amount item type, final test, and warning group, $F(1, 150) = 5.95$, $p = .016$. $\eta_p^2 = .04$. No other effects of interactions were significant, item type: $F = 2.96$, $p = .09$; final test: $F < 1$; warning group: $F < 1$; Item Type × Warning Group: $F < 1$; Final Test × Warning Group: $F < 1$.

To decompose the three-way interaction, we conducted separate 2 (warning group: no warning, warning) × 2 (item type: neutral, misleading) ANOVAs for each of the final test types. When participants took a cued recall test, mean confidence was greater on misleading trials ($M = 71.80$, $SE = 1.83$) as compared with neutral trials ($M = 67.95$, $SE = 1.89$), $F(1, 74) = 7.25$, $p = .009$, $\eta_p^2 = .09$. No other effects or interactions were significant, $Fs < 2$. When participants took the MOT we found an interaction between item type and warning group, $F(1, 76) = 4.46$, $p = .038$, $\eta_p^2 = .06$. When participants were not given a warning, mean confidence was greater associated with responses given on neutral trials ($M = 70.15$, $SE = 2.53$) as compared with misleading trials ($M = 66.91$, $SE = 2.40$). However, when participants were warned, mean confidence on misleading trials ($M = 71.96$, $SE = 2.79$) was greater than that found on neutral trials ($M = 69.36$, $SE = 2.89$). No other effects were significant, $Fs < .1$

Discussion

To briefly review the results from Experiment 1, when participants took a cued recall test and did not receive a warning, we found the standard misinformation effect. That is, memory accuracy was lower for misleading trials relative to neutral trials. When participants received a warning, they reduced the difference in accuracy between misleading and neutral trials. Additionally, when participants took the MOT test, accuracy was higher overall. Regarding average confidence, participants who took the cued recall test were more confident in responses on misleading trials as compared with neutral trials. For the MOT test, participants who did not receive the general warning demonstrated higher confidence on neutral relative to misleading trials. In contrast, those who took the MOT and were warned demonstrated higher confidence on misleading trials compared with neutral ones. While this is an atypical finding, it is not totally unexpected in the context of our procedure. First, accuracy was numerically higher for misleading trials in that group. Given that the MOT provides misleading information as part of the test trial, by providing participants with a highly accessible but incorrect item to exclude, participants may have felt more confident in the final answer. In general, these results replicate the work done by Eakin et al. (2003), demonstrating the efficacy of any form of warning for improving accuracy and resisting misinformation in situations where the

misinformation is not highly accessible (the standard misinformation paradigm).

In Experiment 2, we used the procedure from Experiment 1 in the context of the RES methodology where participants take a test immediately after watching the original event and prior to exposure to the misleading information. Based on the findings of Eakin et al. (2003) and the prior research that initial testing influences the accessibility of postevent information, one hypothesis is that warnings would be ineffective in the context of the RES methodology. Alternatively, Thomas et al. (2010) found general warnings reduce RES; therefore, Experiment 2 was designed to assess whether warnings could in fact foster better discrimination between the misinformation and the original event and, even in the context of highly accessible misinformation, reduce misinformation susceptibility.

Experiment 2

Method

Participants

We conducted effect-size sensitivity analyses in the same fashion as the previous experiment. For the results of these analyses, refer to Table 2. Here, we again focus on the two-way, within-between subjects interactions that are most relevant to our hypotheses. For tests of these interactions, our sample was enough to detect an effect size of .010 with a power level of 90% ($\alpha = .05$, $r = .64$). The effect sizes that we observed were larger than the minimum detectable effect size ($\eta_p^2 = .10$ [Item Type × Final Test] and $\eta_p^2 = .05$ [Item Type × Warning Group]).

A total of 189 undergraduate students volunteered to participate in this study for course credit ($M_{age} = 20.00$, $SD_{age} = 3.29$). Students were recruited from Stockton University ($N = 130$; $M_{age} = 20.43$, $SD_{age} = 3.85$) and Tufts University ($N = 59$; $M_{age} = 19.10$, $SD_{age} = 1.05$). Participants were randomly assigned to one of four between-subjects groups: cued recall no warning ($n = 40$; all from Stockton University), cued recall and a warning ($n = 51$; 30 from Stockton University), MOT no warning ($n = 48$; 30 from Stockton University), or MOT with a warning ($n = 50$; 30 from Stockton University).

Design

The same experiment design used in Experiment 1 was used in Experiment 2.

Materials and procedure

The materials used in Experiment 2 were identical to those used in Experiment 1. Similarly, the methodology

of Experiment 2 followed that of Experiment 1, except for one difference. In Experiment 2, immediately after the video, instead of completing a filler task, participants were asked to answer 24 cued recall questions concerning specific details from the video (e.g., The last assailant took a _____ out of a drawer). Each question remained on the screen for 15 s, during which participants could type an answer. After the 15 s elapsed, the screen advanced to the next question. Participants could not advance to the next question before the 15 s time was completed, nor could they revisit prior questions. After answering each question, participants were asked to provide confidence in their answer on a 0 (*complete guess*) to 100% (*completely confident*) scale, and there was no time limit to respond. Questions were presented in the same order as the corresponding details appeared in the video. The time to complete this initial test took on average the same time participants engaged in the filler task used in Experiment 1.

After this initial test, the procedure followed that described for Experiment 1. That is, participants were presented with the postevent narrative, and depending on group, were given a general warning or no warning. Then they took either the cued recall or MOT test. The final test procedure was identical to that used in Experiment 1. The questions on the initial test were identical to the final cued recall test.

Results

Initial test performance

Performance on the cued recall test that immediately followed the presentation of the video did not differ as a function of item type, warning group, or final test ($M = .60$, $SE = .03$), Fs < 1.

Final test accuracy

We conducted a 2 (final test: cued recall, MOT) × 2 warning group: no warning, warning) × 2 (item type: neutral, misleading) mixed design ANOVA on average proportion correct on the final test of memory. Consistent with the large body of research demonstrating RES, we found a main effect of item type, $F(1, 185) = 45.87$, $p < .001$, $\eta_p^2 = .20$. Participants were more accurate on neutral trials ($M = .66$, $SE = .02$) as compared with misleading trials ($M = .57$, $SE = .01$). We also found a main effect of warning group, $F(1, 185) = 7.39$, $p = .007$, $\eta_p^2 = .04$. People who received a warning ($M = .65$, $SE = .02$) were more accurate than those who were not warned ($M = .58$, $SE = .02$). Finally, we found a main effect of test type, $F(1, 185) = 15.78$, $p < .001$,
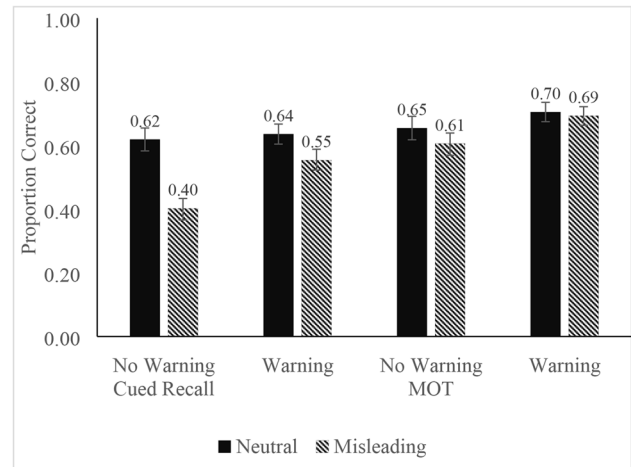

Fig. 3 Accuracy on the final test as a function of item type, warning group, and type of final test for Experiment 2. (*M* and *SE* plotted)

$\eta_p^2 = .08$. People who took the MOT test ($M = .67$, $SE = .02$) were more accurate than those who took the cued recall test ($M = .56$, $SE = .02$). Although we did not find an interaction across all three variables, we did find that item type interacted with warning group, $F(1, 185) = 10.54$, $p < .001$, $\eta_p^2 = .05$, and item type interacted with final test, $F(1, 185) = 20.22$, $p < .001$, $\eta_p^2 = .10$. Figure 3 illustrates the warning has greater effects on misleading trials, $t(187) = 3.08$, $p = .002$, $d = .45$, as compared with neutral trials, $t(187) = 1.11$, $p = .271$. This pattern mirrors the results found in Experiment 1, except for performance on the cued recall test when participants were unwarned. Given that this is the set of conditions under which *retrieval enhanced suggestibility* is most likely to occur, this difference is not surprising. We will return to this point later in our analysis of RES.

Withholding

Importantly, as with Experiment 1, the likelihood of withholding responses was low ($M = .10$); however, to determine whether withholding was impacted by warning or type of final test, we conducted a 2 (final test: cued recall, MOT) × 2 (warning group: no warning, warning) × 2 (item type: neutral, misleading) mixed-design ANOVA on average proportion of withholding responses on the final test. We found an interaction across all variables, $F(1, 185) = 5.07$, $p = .026$, $\eta_p^2 = .10$. No other effects were significant, Fs < 1.

To decompose the three-way interaction, we conducted separate 2 (warning group: no warning, warning) × 2 (item type: neutral, misleading) ANOVAs for each of the final test types. For participants who took a cued recall final test, we found a main effect of item type, $F(1, 89) = 7.09$, $p = .009$,

$\eta_p^2 = .07$. We also found a Warning Group × Item Type interaction, $F(2, 89) = 5.12$, $p = .026$, $\eta_p^2 = .05$. Although withholding did not statistically vary on neutral trials as a function of warning ($M = .07$, $SE = .02$) or no warning ($M = .08$, $SE = .01$), participants were more likely to withhold answers on misleading trials if they had been warned ($M = .06$, $SE = .01$) as compared with if they had not been warned ($M = .02$, $SE = .01$). Alternatively, when participants took the MOT, we found a main effect of item type, $F(1, 96) = 5.80$, $p = .018$, $\eta_p^2 = .06$, but no interaction, $F = 1.2$. Not surprisingly, participants were more likely to withhold answers on misleading trials ($M = .16$, $SE = .02$) as compared with neutral trials ($M = .12$, $SE = .02$).

Confidence

As with Experiment 1, warning also had an impact on average confidence ratings. We conducted a 2 (final test: cued recall, MOT) × 2 (warning group: no warning, warning) × 2 (item type: neutral, misleading) mixed-design ANOVA on average confidence associated with answers on the final test. We found an interaction amount item type and final test, $F(1, 185) = 7.15$, $p = .008$, $\eta_p^2 = .04$. No other effects were significant, $F < 2$. When participants took the MOT, their average confidence was higher on neutral trials ($M = 70.76$, $SE = 1.60$) as compared with misleading trials ($M = 68.77$, $SE = 1.84$). However, the reverse was true when participants took a cued recall test. In this case participants demonstrated higher confidence on misleading trials ($M = 71.06$, $SE = 1.49$) as compared with neutral trials ($M = 68.44$, $SE = 1.65$). The full breakdown of average confidence in Table 3.

Comparison between experiments to examine retrieval enhanced suggestibility

Although directly testing for RES was not the aim of these experiments, we did conduct an analysis on memory accuracy *and* production of misinformation comparing the standard misinformation group from Experiment 1 to the repeated test group from Experiment 2 in order to confirm the occurrence of RES. Importantly, this comparison was restricted to participants in the no warning groups who had taken a cued recall final test. We computed a 2 (number of tests: one [standard], two [repeated]) × 2 (item type: neutral, misleading) mixed-design ANOVA on final test accuracy, and found a main effect of item type, $F(1, 80) = 54.01$, $p < .001$, $\eta_p^2 = .40$. Participants were less accurate on misleading trials as compared with neutral trials. In addition, we found evidence for RES by observing an interaction between item type and number of tests, $F(1, 80) = 6.19$, $p = .015$, $\eta_p^2 = .07$. This pattern is illustrated in Fig. 4.

Finally, we examined whether the production of misinformation on the final cued recall test differed between

Table 3 Mean confidence on the final test (standard error in parentheses)

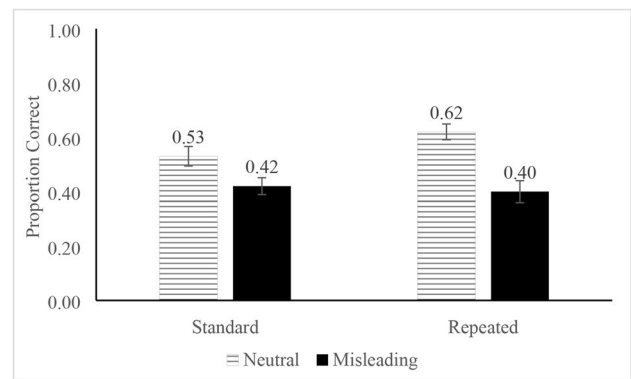|	No warning		Warning	
	Neutral	Misleading	Neutral	Misleading
Experiment 1 Standard				
CR	67.42 (2.87)	72.88 (2.67)	68.62 (2.36)	70.47 (2.45)
MOT	70.15 (2.54)	66.92 (2.40)	69.36 (2.89)	71.96 (2.75)
Experiment 2 RES				
CR	70.60 (2.17)	73.07 (2.09)	66.75 (2.40)	69.49 (2.08)
MOT	72.28 (2.04)	69.03 (2.83)	69.31 (2.45)	68.51 (2.42)

Fig. 4 Proportion correct on the final test comparing the standard to repeated testing for unwarned participants who took a cued recall test. (*M* and *SE* plotted)

participants in the standard misinformation group of Experiment 1 and the repeated test group of Experiment 2. Again, we restricted our comparison to participants who only took the cued recall test and were not warned. We found that participants in the repeated test group were far more likely to have produced misinformation on misleading trials ($M = .44$, $SE = .04$) as compared with participants in the standard misinformation group ($M = .29$, $SE = .02$), $t(80) = 3.17$, $p = .002$, $d = .65$. RES was initially conceptualized by Chan et al. (2009) and named by Thomas et al. (2010) and is understood as changes in accuracy on the final test and/or an increase in production of misleading details. Evidence of RES traditionally is established by the interaction between item type and number of tests and/or a change in average production of misleading details as a function of number of tests taken. In the present data, we provided clear evidence of RES as measured by misinformation production. This approach to defining RES has been used in numerous manuscripts that have reported similar patterns of results in accuracy and misinformation production (e.g., Chan & Langley, 2011; Chan et al., 2009; Experiment 1b; Chan, Wilford, & Hughes, 2012; Chan et al., 2021; Gordon et al., 2015; Manley & Chan, 2019; Wilford et al., 2014).

General discussion

The goal of the present research was to evaluate how different types of warnings affected misinformation with differing levels of accessibility. We investigated the effect of different types of warnings on low accessibility misinformation (the standard misinformation paradigm) and high accessibility misinformation (the RES paradigm). In Experiment 1, we found that both the general warning and the MOT increased accuracy on misleading trials. When participants were not warned, they were significantly less accurate on misleading trials as compared with neutral trials. This generally replicates the findings by Eakin et al. (2003). Specifically, when participants were given any warning, their accuracy improved on misleading trials. Critically, however, these results correspond to the low accessibility conditions in Eakin et al. (i.e., when the misinformation was presented a single time in the context of a narrative).

In Experiment 2, both the general warning and the MOT improved accuracy on misleading trials when repeated testing was implemented. However, unlike in Experiment 1, when participants received a general warning, they continued to demonstrate the misinformation effect as measured by the difference in accuracy between neutral and misleading trials. Importantly, the difference between neutral and misleading items was still greatly reduced in the general warning condition relative to no warning (9% vs. 22%). Experiment 2 used the RES methodology, which, we have argued, leads to highly accessible misinformation. Although Eakin et al. (2003) found that warnings did not reduce misinformation susceptibility for highly accessible misinformation, we found that warnings were effective in this experiment. That said, similar to Eakin et al., the general warning was only moderately effective, with participants still demonstrating misinformation susceptibility, albeit to a lesser degree than when not warned.

Given the discrepancy between Eakin's findings and the present results of Experiment 2 (see also Thomas et al., 2010) we suggest that warnings in the context of the RES paradigm may result in improved contextual discrimination between two highly accessible sources of information (original event memory and postevent narrative). Research suggests that retrieval will increase the accessibility of retrieved information (original event details), and also impact contextual separation between the original tested information and information presented after the test (cf. Whiffen & Karpicke, 2017). Chan and McDermott (2007) demonstrated that source judgments in a list discrimination paradigm were more accurate after testing compared with when participants only studied the information. Specifically, participants either studied word lists or were given free recall tests on these lists prior to a final recognition test with a source judgment. Engaging in a source monitoring task, while not a warning about the veracity of specific content, likely requires participants to engage in retrieval monitoring processes that result in increased scrutiny of responses. Similarly, Gordon and Thomas (2014) demonstrated that participants were better able to remember original event and postevent details in the context of a repeated testing as compared with a single testing paradigm. The strength of the memory representations for the original and postevent information, when paired with certain warnings that encourage participants to more closely inspect their responses, is one potential mechanism that underlies the increased accuracy in the present RES results. In the absence of the warnings, or when warnings are more general, the increased accessibility of the misleading information results in the typical RES effect (i.e., larger misinformation effects).

Confidence and warnings

When examining the confidence, we found across both experiments, that when participants took a cued recall test, responses on misleading trials accompanied higher confidence than on neutral trials. This pattern is consistent with earlier findings using both the standard eyewitness misinformation methodology (e.g., Weingardt et al., 1994), and RES misinformation methodology (e.g., Thomas et al., 2010). In contrast, when participants took the MOT, they generally demonstrated higher confidence on neutral trials as compared with misleading. This was found across both experiments, with one exception.

In Experiment 1, when participants took the MOT and were warned, they showed higher confidence on misleading as compared with neutral trials. One possible reason this counterintuitive pattern emerged is that presentation of the misleading detail in the context of the MOT may have supported a more careful search of their memory for the correct video information. Previous research (Koriat, 1993; Thomas et al., 2012) has demonstrated that retrieval-based metacognitive judgments are based (at least partially) on accessible contextual information, and more information becomes available with more careful searches of memory. The warning may have aided in suppressing this incorrect information, and it follows that confidence may well increase for the remaining candidate response. Thus, pairing the MOT with a warning may have resulted in higher confidence in retrieved responses.

In contrast, when participants took the MOT in Experiment 2, they may have been less able to discount the subjective experience associated with the higher accessibility of the misleading information which resulted in lower confidence on the misleading trials. That said, this is a post hoc

explanation and more targeted investigation would need to be conducted to provide support for these possibilities.

Conclusions, limitations, and future directions

The present study was conducted to better understand whether the accessibility of misleading information affected the efficacy of different types of warnings. We examined both general warnings and warnings that provided specific cautionary information (in the form of the MOT). In Experiment 1, we found that warnings were associated with improvements in accuracy on misleading trials, replicating work done by Eakin et al. (2003). In Experiment 2, we replicated the standard RES effect (increased susceptibility to misleading information as a result of initial testing) in the no warning cued recall group and demonstrated that the effects of even highly accessible misleading information could be reduced via warnings. Taken together, these results suggest that both general and item-level warnings are effective in reducing misinformation susceptibility for both low and highly accessible misinformation. That said, the efficacy of warnings may not generalize to other conditions of highly accessible misinformation. As we have argued, given that we increased misinformation accessibility by requiring retrieval prior to misinformation presentation, we likely also increased accessibility of original event details (cf. Roediger, III & Karpicke, 2006). Thus, situations where misinformation accessibility is increased without strengthening the memory for the original event may not yield the same pattern of results. Direct comparisons of RES with other methods of increasing misinformation accessibility (repetition, for example) would be useful to determine the efficacy of warnings in other situations where misleading information is highly accessible.

One primary limitation of the current research is that it is difficult to compare the magnitude of the differences between warning conditions across experiments. We did demonstrate the standard pattern of RES between Experiments 1 and 2, but a finer grained analysis of specific warning conditions was precluded by statistical power considerations. While we can draw the reader's attention to numerical patterns in the data (better accuracy for the combined warnings), any firm conclusions would be speculative.

From the perspective of real-world eyewitness situations, implementing effective warnings still represents significant challenges. With regards to the MOT, it is unlikely that any members of the criminal justice system would be aware of what postevent misleading information to which a witness had been exposed. However, general warnings that encouraged participants to more closely inspect their retrieved responses, improve both accuracy and reduced the overall confidence in misleading responses. Previous research (Bulevich & Thomas, 2012) demonstrated that warnings that

provided information on what types of contextual information were associated with different sources could improve misinformation accuracy.

Initial testing of an eyewitness event (RES) remains a double-edged sword when exposure to misleading information is possible. An initial test will increase witnesses' memory for the original event but has the possibility of making any subsequent misleading postevent information more accessible. Our research suggests that in situations where witnesses will be recalling information on multiple occasions, those repeated retrievals should be paired with instructions that witnesses should engage in a careful search of their memory and focus on relevant source specifying cues.

Appendix

Instructions for narrative:

In this part of the experiment, you will be presented with a narrative of the episode you just watched. You are instructed to both listen to the narrative and read the narrative. The narrative will be presented over the headphones. You should follow along with the aural presentation using the printed packets at your station. The written packet is divided into sections that correspond to the auditory presentation. It is important that your pace of reading matches the aural presentation. The experiment will walk around to make sure that the pace of reading matches the aural presentation. Please listen and read the narrative very carefully, as it may help refresh your memory for the memory test.

Please ask any questions now. Otherwise, press SPACE to continue.

Warning:

You will have to answer questions regarding the episode you previously watched. We just presented a narrative of that episode; however, we are uncertain as to the source of the narrative. Therefore, we were unable to verify the accuracy of the narrative. As such, base your answer only on what you saw in the VIDEO, and not on what was presented in the narrative.

Instructions for final test:

A short time ago, you watched a video. Now, you will answer some questions about the video.

Please answer the following questions to your best ability. Be as accurate as possible, and answer as many questions as you can. Respond only with information you remember being presented in the VIDEO.

Each question will require only a brief answer. You will have 15 seconds to answer each question. There is no need to press "ENTER." The next question will show up automatically.

Following each question, you answer, you will be asked to rate your confidence in the answer you provided. Please type in a value ranging from 0 (*completely unconfident*) to 100 (*completely confident it was in the VIDEO*). An answer you are completely sure of would be given a 100, while a complete guess would be given a 0.

If you have any questions, please ask your experimenter now. Otherwise, please let the experimenter know that you are ready to start.

References

Belli, R. F., Windschitl, P. D., McCarthy, T. T., & Winfrey, S. E. (1992). Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 356–367. https://doi.org/10.1037/0278-7393.18.2.356

Belli, R. F., Lindsay, D. S., Gales, M. S., & McCarthy, T. T. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. *Memory & Cognition, 22*(1), 40–54. https://doi.org/10.3758/BF03202760

Bezard, R., Bérard, H., Cabaud, P., & Dassin, J. (1955). *Rififi* [Motion picture]. France, Pathé Consortium Cinéma.

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of postwarning studies. *Journal of Applied Research in Memory and Cognition, 3*(2), 77–88. https://doi.org/10.1016/j.jarmac.2014.03.005

Brackmann, N., Otgaar, H., Sauerland, M., & Howe, M. L. (2016). The impact of testing on the formation of children's and adults' false memories. *Applied Cognitive Psychology, 30*(5), 785–794. https://doi.org/10.1002/acp.3254

Bulevich, J. B., & Thomas, A. K. (2012). Retrieval effort improves memory and metamemory in the face of misinformation. *Journal of Memory and Language, 67*(1), 45–58. https://doi.org/10.1016/j.jml.2011.12.012

Butler, B. J., & Loftus, E. F. (2018). Discrepancy detection in the retrieval-enhanced suggestibility paradigm. *Memory*, *26*(4), 483–492. https://doi.org/10.1080/09658211.2017.1371193

Chan, J. C. K., & Langley, M. M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 248–255. https://doi.org/10.1037/a0021204

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 431–437. https://doi.org/10.1037/0278-7393.33.2.431

Chan, J. C., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological science, 20*(1), 66–73. https://doi.org/10.1111/j.1467-9280.2008.02245.x

Chan, J. C. K., Manley, K. D., & Lang, K. (2017). Retrieval-enhanced suggestibility: A retrospective and a new investigation. *Journal of Applied Research in Memory and Cognition, 6*(3), 213–229. https://doi.org/10.1016/j.jarmac.2017.07.003

Chan, J. C. K., O'Donnell, R., & Manley, K. (2021). Warning weakens retrieval-enhanced suggestibility only when it is given shortly after misinformation: the critical importance of timing. *Journal of Experimental Psychology Applied*. https://doi.org/10.1037/xap0000394

Cohen, J (1988). *Statistical power analysis for the social sciences* (2nd ed.). Erlbaum.

Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(5), 813–825. https://doi.org/10.1037/0278-7393.29.5.813

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

E-Prime (Version 2.1). (n.d.). [Computer software]. Psychology Software Tools.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science, 20*(1), 20–23. https://doi.org/10.1177/0963721410396620

Gallo, D. A. (2004). Using recall to reduce false recognition: diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 120–128. https://doi.org/10.1037/0278-7393.30.1.120

Giner-Sorolla, R., Aberson, C. L., Bostyn, D. H., Carpenter, T., Conrique, B. G., Lewis, N. A., & Soderberg, C. (2019). *Power to detect what? Considerations for planning and evaluating sample size*. Advance online publication. https://osf.io/jnmya/

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

Gordon, L. T., & Thomas, A. K. (2017). The forward effects of testing on eyewitness memory: The tension between suggestibility and learning. *Journal of Memory and Language, 95,* 190–199. https://doi.org/10.1016/j.jml.2017.04.004

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

Gordon, L. T., Bilolikar, V. K., Hodhod, T., & Thomas, A. K. (2020). How prior testing impacts misinformation processing: A dual-task approach. *Memory & Cognition, 48*(2), 314–324.

Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning & Verbal Behavior, 21*(2), 207–219. https://doi.org/10.1016/S0022-5371(82)90571-0

Higham, P. A., Blank, H., & Luna, K. (2017). Effects of postwarning specificity on memory performance and confidence in the eyewitness misinformation paradigm. *Journal of Experimental Psychology: Applied, 23*(4), 417–432. https://doi.org/10.1037/xap0000140

Karanian, J. M., Rabb, N., Wulff, A. N., Torrance, M. G., Thomas, A. K., & Race, E. (2020). Protecting memory from misinformation: Warnings modulate cortical reinstatement during memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(37), 22771–22779. https://doi.org/10.1073/pnas.2008595117

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609

Lakens, D. (2013 Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, 12.

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

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6), 1077–1083. https://doi.org/10.1037/0278-7393.16.6.1077

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17*(3), 349–358. https://doi.org/10.3758/BF03198473

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*(1), 19–31. https://doi.org/10.1037/0278-7393.4.1.19

Loftus, E. F., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition, 17*(5), 607–616. https://doi.org/10.3758/BF03197083

Manley, K. D., & Chan, J. C. K. (2019). Does retrieval enhance suggestibility because it increases perceived credibility of the postevent information? *Journal of Applied Research in Memory and Cognition, 8*(3), 355–366. https://doi.org/10.1016/j.jarmac.2019.06.001

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

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

Rindal, E. J., DeFranco, R. M., Rich, P. R., & Zaragoza, M. S. (2016). Does reactivating a witnessed memory increase its susceptibility to impairment by subsequent misinformation? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(10), 1544–1558. https://doi.org/10.1037/xlm0000265

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Tools.

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*(2), 149–157. https://doi.org/10.1016/j.jml.2010.04.004

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2012). An analysis of the determinants of the feeling of knowing. *Consciousness and Cognition: An International Journal, 21*(4), 1681–1694. https://doi.org/10.1016/j.concog.2012.09.005

Thomas, A. K., Gordon, L. T., Cernasov, P. M., & Bulevich, J. B. (2017). The effect of testing can increase or decrease misinformation susceptibility depending on the retention interval. *Cognitive Research: Principles and Implications, 2,* 45. https://doi.org/10.1186/s41235-017-0081-4

Weingardt, K. R., Toland, H. K., & Loftus, E. F. (1994). Reports of suggested memories: Do people truly believe them? In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (p. 3–26). Cambridge University Press. https://doi.org/10.1017/CBO9780511759192.002

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1036–1046. https://doi.org/10.1037/xlm0000379

Wilford, M. M., Chan, J. C., & Tuhn, S. J. (2014). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. *Journal of Experimental Psychology: Applied*, *20*(1), 81–93. https://doi.org/10.1037/xap0000001

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.