scientific reports

OPEN

Timing of pre-retrieval warnings matters in reducing memory errors in a repeated testing misinformation study

Alia N. Wulff[1], Jessica Karanian[2], Elizabeth Race[3] & Ayanna K. Thomas[3]✉

Individuals are often exposed to information regarding previously witnessed events. The misinformation effect occurs when inaccurate post-event information impairs accuracy on a subsequent test of memory for the original event. The misinformation effect is increased when a test is given prior to exposure to post-event information, a phenomenon termed Retrieval Enhanced Suggestibility (RES). In two experiments, we investigated the value of general warnings prior to retrieval towards reducing RES, in situations where reconsolidation due to test related activation was likely. In both experiments, participants were exposed to an original event, and after a 24-hour retention interval either took a test about that original event or were not tested before being exposed to a narrative that included misleading details about the original event. In Experiment 1, a warning about the quality of the post-event narrative was then given followed by the final test. In Experiment 2, the warning and final test were delayed by 24 h. We found that warnings were effective in Experiment 1 in reducing RES, suggesting that even when initial testing supported learning of PEI, original details remained accessible. However, in Experiment 2, when warnings were delayed, participants were no longer able to effectively use them to reduce RES. These results suggest that warnings are most effective when given in close temporal proximity to misinformation.

Keywords Eyewitness memory, Retrieval enhanced suggestibility, Interpolated testing, Misinformation effect

Researchers have consistently demonstrated that exposure to post-event information (PEI) may negatively impact memory for the original event[1–3]. This has been termed the eyewitness *misinformation effect* and has been a widely studied phenomenon for decades. More recently, Chan, Thomas, and colleagues have examined the influence of repeated testing on the misinformation effect[4,5]. They introduced a test of the original event prior to exposure to post-event information. Using this methodology, they have consistently demonstrated that taking a test prior to exposure to PEI (interpolated testing) increases susceptibility to that information. This effect has been termed *retrieval enhanced suggestibility* (RES).

Retrieval Enhanced Suggestibility

Researchers have suggested that retrieval of original event details prior to the presentation of PEI may increase the susceptibility of retrieved original details to interference from PEI. According to Chan and LaPaglia[6] initial retrieval places original event memories in a malleable state. When PEI is presented, that information may alter original memories, resulting in new distorted memories[7,8]. Importantly, such RES effects are most prominent in situations where the original event has had time to consolidate prior to retrieval-based activation. In these situations, original event memories are more likely to be updated and modified by new PEI.

However, research also suggests that in some situations original event memories may remain accessible even in the context of having taken an initial test after the original information, and before the presentation of PEI. As one example, using a similar interpolated testing procedure as Chan, Thomas, and colleagues, Rindal[9] did not find RES. Specifically, Rindal used a final two alternative forced choice recognition test where the choices were the original detail and an unrelated lure, and found no evidence for enhanced suggestibility. These results highlight the possibility that taking an initial test may impact how subsequent misinformation may inhibit, or block, retrieval of the original memory, but may not alter the original event memory.

[1]Department of Psychology, Arizona State University, Phoenix, AZ, USA. [2]Department of Psychological and Brain Sciences, Fairfield University, Fairfield, CT, USA. [3]Department of Psychology, Tufts University, Medford, MA, USA. ✉email: ayanna.thomas@tufts.edu

Although Rindal et al.[9] provided compelling evidence against an altered memory hypothesis, altered memories due to interpolated testing may be dependent on reactivation of original event memories prior to the presentation of PEI (cf[6]). Reactivation can only occur once a memory has been consolidated. In the present study, we reactivated the original event after a 24-hour retention interval to allow for consolidation. Following reactivation through testing, we presented the PEI that included details consistent and inconsistent with the original event. In this context, PEI is more likely to disrupt memories associated with original event details. Additionally, PEI may be better learned and more easily retrieved than the original event details because reactivation through testing may change how the PEI is processed, resulting in what has been termed a *forward effect of testing*[10].

Research suggests that interpolated testing may influence how PEI is learned and remembered[5,10–12]. Gordon and Thomas[10] found that participants who were given an interpolated test of memory allocated more attention, as measured by reading time, towards details in the post-event narrative than did participants who were not given an interpolated test of memory. They hypothesized that the increased attentional allocation towards those details improved learning and retention of those details. Therefore, RES may result from both reconsolidation mechanisms, and test-based potentiation effects, where reactivation through testing potentiates learning of new incorrect information. Warnings offer a way to differentiate between these two possibilities.

Warnings

Thomas and colleagues found that when participants were warned about the presence of misleading PEI prior to a final memory test, RES was eliminated[5]. Thomas et al. argued that the final-test accuracy decrements on misleading trials that occurred when participants were given an interpolated test were due to improved learning of PEI, and that warnings helped participants leverage more effortful retrieval processes to access and report original event details.

Pre-retrieval warnings (i.e., warnings given after the PEI but before the final test) provide a useful test of possible mechanisms proposed to account for RES. Research suggests that pre-retrieval warnings may reduce reliance on retrieval fluency and promote careful source discrimination in the standard misinformation experiment (for review see[13]). That is, warnings may encourage the consideration of source monitoring cues associated with original event information and PEI. With more attention to cues that may differentiate these two sources of information, participants may be better able to remember original event details. While pre-retrieval warnings have been shown to reduce RES (e.g[5,14]) these studies examined these effects without considering consolidation and reactivation mechanisms. That is, original event presentation, interpolated testing, PEI, and final testing all occurred in one testing session within a two-hour time block.

The present study investigated whether warnings prior to taking the final test could effectively reduce RES in situations where the original event had time to consolidate prior to reactivation. More recently, Chan, O'Donnell, and Manley[15] replicated the pattern observed by Thomas et al.[5] and found that warnings prior to the final test could be effective at reducing RES even when the final test was delayed by 48 h. That is, warnings were effective in a RES study where the original event was followed by an immediate test, the PEI, and a warning about the PEI. In this context, when the final test occurred 48 h after the warning, the warning remained effective. That said, the original event may not have had sufficient time to consolidate prior to the first test. Research examining the efficacy of warnings when original learning has had time to consolidate provides for a more rigorous test of original memory accessibility.

The present study

The present study examined pre-retrieval warnings in a repeated testing eyewitness experimental design where the original event has had the opportunity to consolidate (Experiment 1) and when both the original event and PEI has had the opportunity to reconsolidate (Experiment 2). Unlike prior studies, a 24-hour retention interval followed witnessing the original event. In addition to allowing for original memory consolidation, the 24-hour retention interval following original event encoding was predicted to reduce the value of retrieval on original event long-term memory (e.g[16], and still exert an influence on learning and incorporating PEI (e.g[17]). Therefore, the initial test was used to reactivate original event details without necessarily supporting long-term memory for those details. We also examined average confidence as research has consistently demonstrated that participants are often highly confident in misinformation[18], and warnings may improve the confidence-accuracy relationship[5,14,15,19].

In Experiment 1, participants were exposed to an original event, and after 24 h, took an initial test or engaged in a distractor task, and then were presented with PEI in the form of a written narrative. After a short delay, half of these participants received a general warning, and then all participants received a final test of memory. In Experiment 2, participants followed a similar procedure except that the warning and final test occurred 24 h after the PEI. We predicted that warning would operate to reduce both the misinformation effect, as measured by a decrement in accuracy on misleading trials, and RES effect, as measured by the interaction between trial type and test condition, in Experiment 1. These results would suggest that original event details remained accessible even when conditions to support alteration through reactivation and re-learning were met. In Experiment 2, we predicted that warnings would be ineffective, in a situation where the initial test of memory was delayed replicating the findings by Chan et al.[15], suggesting that when conditions for memory alteration during reconsolidation are met, warnings cannot assist with original memory access. Further, we predicted that while warnings would reduce average confidence found on misleading trials in Experiment 1, it would have no effect on average confidence when delayed by 24 h (Experiment 2).

Experiment 1

Our goal was to examine the effectiveness of pre-retrieval warnings in reducing RES in a context where testing of the original event was less beneficial to original event memory and was likely to potentiate learning of post-event information (original event --> 24 h --> test --> PEI --> warning --> test compared to original event --> 24 h --> distractor --> PEI --> warning --> test). A schematic representation of the experimental procedure can be found in Fig. 1.

Method

All experimental procedures and protocols were approved by the Tufts University Social, Behavioral, and Educational Research Institutional Review Board (protocol number 1706030). All experiments and methods were carried out in accordance with Tufts University and Federal Regulation guidelines. Informed consent was obtained from all participants.

Participants

A priori power analyses for each planned statistical analysis were completed to determine sample size. We determined our sample size ($N = 132$) through a priori power analysis of the statistical analysis which required the highest sample size to be appropriately powered (interaction for the 3 × 2 ANOVA). For this calculation, we used a medium effect size ($r = .25$) and a power of 0.8. As such, we expected to collect enough data to reach 35 participants in each condition, or 140 in total. All participants were recruited from the participant pool managed by Prolific. Participants were paid $4 for the first session and $6 for the second session, for a total of $10. A total of 161 participants were recruited for this study ($M_{age} = 35.51$, $SD_{age} = 12.05$; 90 women, 68 men, 3 non-binary). All participants were in the United States at the time of the experiment.

<u>Exclusionary criteria</u> There were several planned exclusionary criteria. First, participants were excluded if they took longer than 90 min to complete the second day of the study. Participants were also excluded if they scored less than 10% on the final test, or fewer than two questions correct. Finally, participants were also excluded if they failed either of two attention check questions. The first of these was "How many fatal heart attacks have you suffered from since the start of this survey?". Any response other than "None" would result in exclusion from data analysis. The second attention check question instructed them to type "green" in response to the question "What is your favorite color?". If participants typed anything other than "green", they were excluded from data analysis. After all exclusionary criteria, analyses were run on 122 participants.

Materials

<u>Witnessed event and audio narrative</u> The video utilized in this study is a 22-minute clip from the black and white French film *Rififi* that contained no dialogue. The clip depicted a burglary of a jewelry store.

The audio narrative of the event was a recording of a synopsis of the original event at a normal speaking rate of 135–160 words per minute. Within this narrative there were 24 sentences that included 24 critical details that were included in the memory tests and 91 filler sentences on which participants were not tested. Each critical sentence was separated by at least three filler sentences. The critical sentences contained a consistent detail, a misleading detail, or a neutral detail that did not confirm or disconfirm the original event. Critical details were presented at the end of the sentences. A consistent sentence contained information that matched the details seen in the video (i.e., "The assailants hung a blanket on the window."). A misleading sentence manipulated information seen in

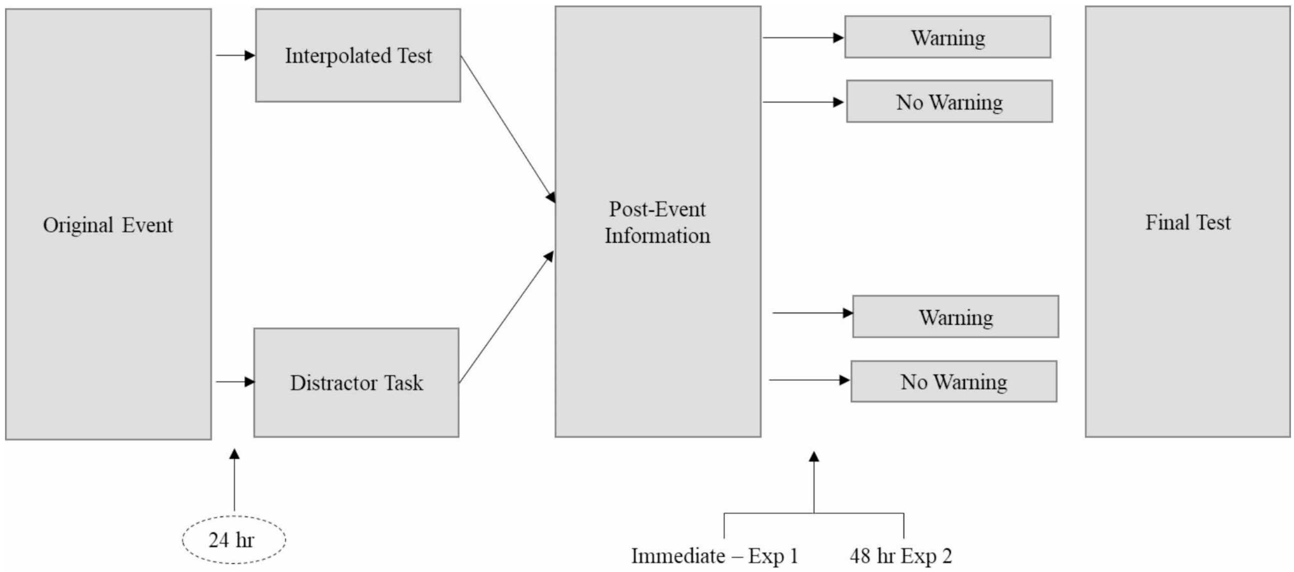

Fig. 1. Methodology flowchart for Experiments 1 and 2.

the video (i.e., "The assailants hung a curtain on the window."). A neutral sentence gave no specific details (i.e., "The assailants blocked the window."). Participants were assigned to one of three counterbalance conditions for the audio narrative. The conditions were counterbalanced using the Latin square method. Each counterbalance contained eight consistent details, eight misleading details, and eight neutral details.

Tests All test questions and misleading details were developed as part of a prior study (cf., Gordon & Thomas, 2017). The questions chosen had an accuracy rate for cued recall between 40 and 70%. Misleading details were developed to be plausible alternatives in the context of the video (e.g., using "white shoes" as the misleading version of "black shoes" instead of an alternative color, as the video clip is in black and white). The initial and final test were identical, as is typical in RES experiments.

All materials were administered on Qualtrics, an online testing software. The test contained 24 forced cued recall questions, directly corresponding to the 24 critical details in the narrative. Participants were told they must provide a response for every question, even if they had to guess, though the testing software itself did not require the participants to enter an answer to proceed to the next question. Participants typed in their answer using their computer keyboard. The questions were asked in the same order they occurred in the video and the same order of questions was given to all participants. After each question participants were asked to provide a confidence rating on a scale from 0 (not confident) to 100 (completely confident). Testing was self-paced and participants could not return to a question once they had proceeded to the next question.

Procedure

Participants first completed consent procedures. Participants then completed a demographics questionnaire and the Shipley Institute of Living Scale: Vocabulary. Their only task for the first experimental session was to watch the video. Participants were instructed to watch the video and pay attention as their memory would be tested as a later point. Participants were given a four-hour window the following day, 22 to 26 h after the first experimental session, to complete thestudy.

When participants began the second session, they first re-read the consent form and confirmed their continued consent to participate. They then responded to an attention check question. Participants then either played a non-verbal computer game for 10 min or were given the interpolated (first) test. They were then given another attention check question. All participants then played a non-verbal computer game for another 10 min. They were then instructed to listen to the audio narrative. After the narrative, participants were either given a warning or not given a warning. When warned participants were told "You will have to answer questions regarding the video you previously watched for a second time. We just played a narrative of that video; 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 you heard in the narrative." All participants were then given the final test. Finally, participants responded to demographic questions and completed a vocabulary test.

Results

Data availability

This research was conducted as part of Alia Wulff's dissertation and was supported by the National Science Foundation Award Number:1728764; Award Title: Using Stress and Warning to Improve Eyewitness Memory. All data are available at https://osf.io/3kf84/?view_only=adc91d71360249a09fcebc56b5ce0438.

Accuracy

Average accuracy, as measured by the proportion correct across all critical questions (e.g., consistent, neutral misleading), for the initial test was 49%. Average accuracy on the initial test serves to indicate what information participants remembered after a 24-hour retention interval. Both the standard misinformation effect and RES are best understood by examining final test performance. Traditionally, the misinformation effect is understood as the decrement in accuracy after the presentation of PEI and is measured by comparing performance on neutral and misleading trials. RES is typically investigated by examining the interaction between trial type and testing group, with the expectation that initial testing will result in worse performance on misleading trials as compared to neutral trials. We first conducted a 2 (testing group: no interpolated test or interpolated test) x 2 (warning group: none or pre-retrieval) x 3 (trial type: consistent, neutral, and misleading) mixed design ANOVA on average accuracy on the final test to investigate both traditional misinformation effects and RES. For the final test, we computed average accuracy as a proportion, based on the number correct out of the total number of questions within a trial type (see Table 1). We found a main effect of trial type, $F(2, 236)=72.46$, $p<.001$, $\eta_p^2=0.38$. All follow-up comparisons were assessed using a Bonferroni adjusted alpha ($\alpha_{Bon}=0.05$/number of tests). Follow-up comparisons ($\alpha_{Bon}=0.017$) found that consistent trials ($M=0.63$) had higher accuracy than neutral trials ($M=0.45$), $t(121)=6.86$, $p<.001$, $d=0.62$, and misleading trials ($M=0.35$), $t(121)=9.56$, $p<.001$, $d=0.87$. Neutral trials had higher accuracy than misleading trials, $t(121)=5.13$, $p<.001$, $d=0.46$, demonstrating the traditional misinformation effect.

We also found an interaction between trial type and testing group, $F(2, 236)=21.35$, $p<.001$, $\eta_p^2=0.15$. Follow-up tests ($\alpha_{Bon}=0.017$) found that participants who were given an iterpolated test were more accurate on consistent trials ($M=0.73$) compared to participants who were not ($M=0.53$), $t(120)=4.73$, $p<.001$, $d=0.86$. Conversely, participants who were given an interpolated test were less accurate on misleading trials ($M=0.30$) than were participants who were not ($M=0.40$), $t(120) = -2.57$, $p=.006$, $d = -0.47$, demonstrating RES. There were no differences between testing groups in accuracy for neutral trials.

Finally, we found a three-way interaction among the manipulated variables, $F(2, 236)=5.19$, $p=.006$, $\eta_p^2=0.04$. As the purpose of this study was to investigate the impact of warning on different testing groups, we followed up on this interaction with two 2 (warning group: none or pre-retrieval) x 3 (trial type: consistent,

	Consistent	Neutral	Misleading	Production
No interpolated test				
No warning	0.47 (0.043)	0.40 (0.040)	0.38 (0.039)	0.26 (0.037)
Warning	0.58 (0.043)	0.48 (0.040)	0.42 (0.039)	0.22 (0.037)
Interpolated test				
No warning	0.79 (0.048)	0.50 (0.039)	0.23 (0.038)	0.58 (0.036)
Warning	0.68 (0.040)	0.44 (0.037)	0.36 (0.036)	0.38 (0.034)

Table 1. Experiment 1 average accuracy and misinformation production on the final test by testing group, warning type, and trial type. Note. Standard errors are in parentheses.

neutral, or misleading) mixed-design ANOVA for each of the testing groups. For no interpolated test group, we found a main effect of trial type, $F(2, 112) = 7.76$, $p < .001$, $\eta_p^2 = 0.12$. Follow-up tests ($\alpha_{Bon} = 0.017$) found that participants were more accurate on consistent trials overall ($M = 0.53$) than on both neutral ($M = 0.44$), $t(57) = 2.46$, $p = .008$, $d = 0.32$, and misleading trials ($M = 0.40$), $t(57) = 3.57$, $p < .001$, $d = 0.47$. There was no significant difference in accuracy between neutral and misleading trials. Warnings did not significantly affect performance for this group.

Amongst participants who were given an interpolated test, we found a main effect of trial type, $F(2, 124) = 85.82$, $p < .001$, $\eta_p^2 = 0.58$. Follow-up tests ($\alpha_{Bon} = 0.017$) found that participants were more accurate on consistent trials overall ($M = 0.73$) than on both neutral ($M = 0.47$), $t(63) = 7.48$, $p < .001$, $d = 0.94$, and misleading trials ($M = 0.30$), $t(63) = 10.96$, $p < .001$, $d = 1.37$. Participants were also more accurate on neutral trials than on misleading trials, $t(63) = 5.67$, $p < .001$, $d = 0.71$. We also found an interaction between trial type and warning, $F(2, 124) = 6.15$, $p = .003$, $\eta_p^2 = 0.09$. Follow-up tests ($\alpha_{Bon} = 0.017$) found that participants were more accurate on misleading trials when warned ($M = 0.36$, $SE = 0.04$) than when unwarned ($M = 0.23$, $SE = 0.04$), $t(62) = 2.18$, $p = .016$, $d = 0.55$.

Misinformation production

Misinformation production, or the production of misleading details on misleading trials when participants fail to remember original event details, was calculated as a proportion. The proportion was based on the number of produced misleading details out of the total number of misleading trials. We conducted a 2 (testing group: no interpolated test or interpolated test) x 2 (warning group: none or pre-retrieval) ANOVA on the average rate of production of misleading information. We found a main effect of warning, $F(1, 118) = 10.55$, $p = .002$, $\eta_p^2 = 0.08$, such that participants who were warned had lower production rates ($M = 0.30$) than participants who were not warned ($M = 0.42$). We also found a main effect of testing, $F(1, 118) = 44.40$, $p < .001$, $\eta_p^2 = 0.27$, such that participants who were not immediately tested had lower rates of production ($M = 0.24$) than participants in the interpolated testing group ($M = 0.48$). We also found an interaction between testing and warning, $F(1, 118) = 5.28$, $p = .023$, $\eta_p^2 = 0.043$. Follow-up tests found that participants who were given an interpolated test produced fewer misleading details when warned ($M = 0.38$) than when unwarned ($M = 0.58$), $t(62) = 3.42$, $p < .001$, $d = 0.86$.

Average confidence

We conducted a 2 (testing group: no interpolated test or interpolated test) x 2 (warning group: none or pre-retrieval) x 3 (trial type: consistent, neutral, and misleading) mixed design ANOVA on average confidence judgments associated with final test performance. We found a main effect of trial type $F(2, 236) = 31.05$, $p < .001$, $\eta_p^2 = 0.21$. Average confidence on consistent trials ($M = 66.27$) was statistically greater than neutral trials ($M = 54.88$), $t(121) = 7.53$, $p < .001$, and misleading trials ($M = 62.00$), $t(121) = 3.36$, $p < .001$. Average confidence on neutral trials was also greater than misleading trials, $t(121) = 4.38$, $p < .001$. We also found a main effect for testing group, $F(1, 118) = 9.98$, $p < .005$, $\eta_p^2 = 0.08$. Average confidence was greater when participants had taken an interpolated test ($M = 66.29$) compared to when they had not ($M = 55.63$). Importantly, we also found an interaction between testing group and warning, $F(1, 118) = 5.80$, $p < .05$, $\eta_p^2 = 0.47$. Consistent with previous research, when participants received an interpolated test, they demonstrated greater confidence on the final test than when they had not. However, a warning reduced the inflated confidence demonstrated in the interpolated testing group. This pattern of results can be seen in Fig. 2.

Discussion

When participants took an initial test, they were less accurate on misleading trials and more likely to produce misleading information on the final test, demonstrating RES. We also found that when participants took an initial test, they were more confident in their final test answers than when they had not taken that test. Importantly, when a warning was given, accuracy on misleading trials increased, misinformation production decreased, and average confidence in responses was reduced. These results suggest that original event information was not altered by reactivation and presentation of PEI. That said, Experiment 1 did not allow for reconsolidation following the presentation of PEI.

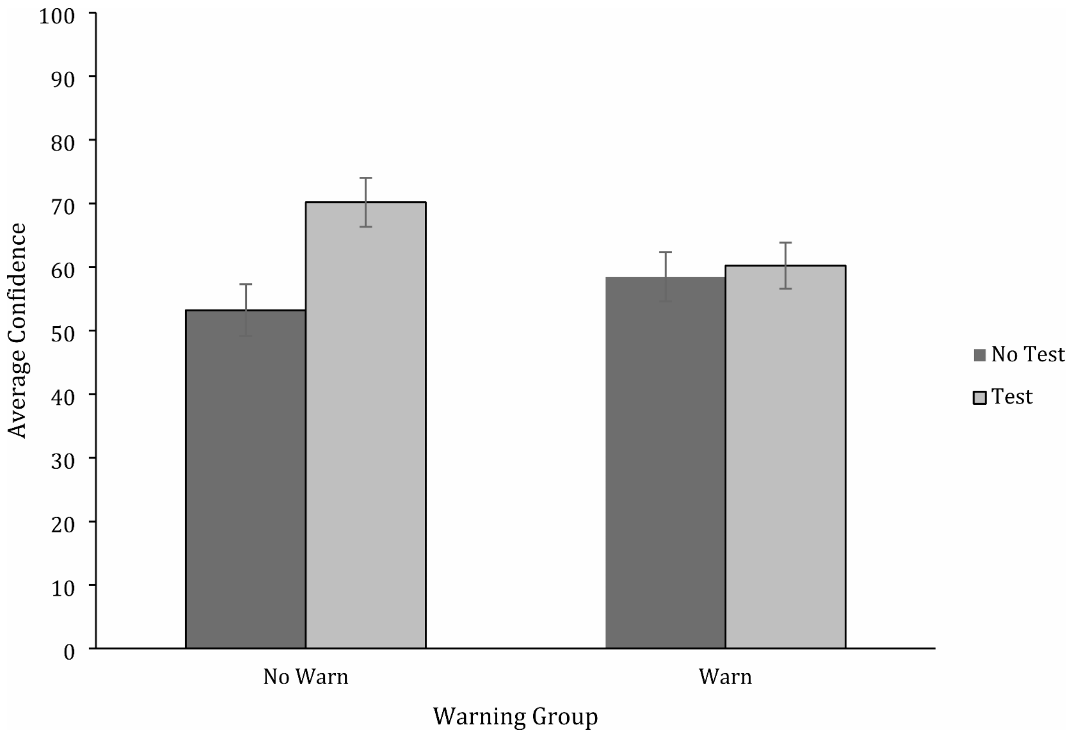

Fig. 2. The relationship between warning and interpolated testing in experiment 1.

Experiment 2

Experiment 2 was designed to test whether warnings could be effective when conditions designed to promote original memory alteration through reactivation and memory updating processes were met. Therefore, participants watched a video on the first day, took an initial test and were exposed to the post-event information on the second day, and then completed warning procedures and the final test on the third and final day of the experiment (original event --> 24 h --> **test** --> PEI --> 24 h --> warning --> test compared to original event --> 24 h --> distractor --> PEI --> 24 h --> warning --> test).

Method

Participants

Using the same values from the power analysis from the previous experiment, we expected to collect enough data to reach 35 participants in each condition or 140 total. All participants were recruited from the participant pool managed by Prolific. Participants were paid $4 for the first session, $4 for the second, and $2 for the final session, for $10. All exclusionary criteria were the same as outlined in the exclusionary criteria for Experiment 1, apart from total time taken in the study. Participants in this experiment were excluded if they took longer than 50 min on the second day or longer than 25 min on the final day of the experiment. Data were collected from a total of 172 participants (M_{age} = 38.65; 79 women, 84 men, 4 non-binary, 5 declined to respond). After all exclusionary criteria, 123 participants were used in analyses.

Materials and procedure

This experiment used the same video, audio narrative, and tests as used in Experiment 1. It was also administered online using Prolific software. This experiment used the same procedure as used in Experiment 1, apart from one deviation. In this experiment, a second 24-hour delay was inserted between the presentation of the post-event information and the final test. Participants first watched the video and then experienced a 24-hour delay. They then took an interpolated test or did not take an interpolated test and were exposed to post-event information before another 24-hour delay. Finally, they either received a warning or did not receive a warning, before taking the final test.

Results

Accuracy

Average accuracy for the interpolated test was 52%. All follow-up comparisons were assessed using a Bonferroni adjusted alpha (α_{Bon} = 0.05/number of tests). We first conducted a 2 (testing group: no interpolated test or interpolated test) x 2 (warning group: none or pre-retrieval) x 3 (trial type: consistent, neutral, and misleading) mixed design ANOVA on average accuracy on the final test. The average accuracy and production on the final test for all conditions can be found in Table 2. We found a main effect of trial type, $F(2, 238) = 54.71$, $p < .001$, $\eta_p^2 = 0.31$. Follow-up comparisons ($\alpha_{Bon} = 0.017$) found that participants were more accurate on consistent ($M = 0.65$, $SE = 0.02$) than neutral ($M = 0.51$, $SE = 0.02$), $t(122) = 6.77$, $p < .001$, $d = 0.61$, and misleading trials

	Consistent	Neutral	Misleading	Production
No interpolated test				
No warning	0.55 (0.048)	0.49 (0.041)	0.42 (0.044)	0.21 (0.044)
Warning	0.61 (0.044)	0.51 (0.038)	0.49 (0.041)	0.20 (0.040)
Interpolated test				
No warning	0.71 (0.040)	0.50 (0.034)	0.32 (0.037)	0.51 (0.037)
Warning	0.73 (0.054)	0.55 (0.046)	0.37 (0.050)	0.37 (0.049)

Table 2. Experiment 2. Average accuracy and misinformation production on the final test by testing group, warning type, and trial type. Note. Standard errors are in parentheses.

	Consistent	Neutral	Misleading
Experiment 1			
No interpolated test			
No warning	56.40 (4.05)	50.00 (4.06)	51.39 (4.29)
Warning	61.36 (3.33)	55.55 (3.72)	59.04 (3.71)
Interpolated test			
No warning	80.27 (3.24)	60.08 (4.48)	73.82 (3.73)
Warning	66.54 (3.48)	53.87 (3.48)	63.15 (3.12)
Experiment 2			
No interpolated test			
No warning	53.74 (4.65)	46.34 (4.10)	50.13 (4.03)
Warning	55.93 (4.39)	46.93 (3.73)	48.88 (3.33)
Interpolated test			
No warning	68.23 (4.87)	50.64 (3.46)	62.66 (3.33)
Warning	61.23 (4.88)	49.20 (4.67)	53.47 (4.39)

Table 3. Average confidence as a function of trial type, testing group, and warning group for experiments 1 and 2. Note. Standard errors are in parentheses.

($M = 0.39$, $SE = 0.02$), $t(122) = 8.74$, $p < .001$, $d = 0.79$. Accuracy was higher on neutral as compared to misleading trials, $t(122) = 4.92$, $p < .001$, $d = 0.63$.

We also found an interaction between trial type and testing group, $F(2, 238) = 13.08$, $p < .001$, $\eta_p^2 = 0.10$. Follow-up tests ($\alpha_{Bon} = 0.008$) found that participants in the no interpolated testing group were more accurate on consistent trials ($M = 0.59$) than on neutral trials ($M = 0.50$), $t(60) = 3.15$, $p = .001$, $d = 0.40$, and on misleading trials ($M = 0.46$), $t(60) = 3.84$, $p < .001$, $d = 0.75$. There were no differences between neutral and misleading trials amongst participants who were not immediately tested. That is, a misinformation effect was not found in this context. However, participants who were tested exhibited differences among all trial types. Consistent trials ($M = 0.71$) had higher accuracy than neutral trials ($M = 0.52$), $t(61) = 6.48$, $p < .001$, $d = 0.82$, and misleading trials ($M = 0.33$), $t(61) = 8.99$, $p < .001$, $d = 1.14$. Neutral trials had higher accuracy than misleading trials, $t(61) = 5.54$, $p < .001$, $d = 0.70$. There were no other main effects or interactions for accuracy on the final test. Importantly, we did not find any impact of warning on final test accuracy.

Production
To investigate rates of production, we conducted a 2 (testing group: no interpolated test or interpolated test) x 2 (warning group: none or pre-retrieval) ANOVA on the average rate of production of misleading information. We found a main effect of testing, $F(1, 119) = 28.83$, $p < .001$, $\eta_p^2 = 0.19$, such that participants who were not immediately tested had lower rates of production ($M = 0.21$) than did participants in the interpolated testing group ($M = 0.44$). Importantly, we found no effect of warning on production of misleading details, and no interaction between testing and warning.

Average confidence
Average confidence judgements can be found in Table 3. We conducted a 2 (testing group: no initial test or initial test) x 2 (warning group: none or pre-retrieval) x 3 (trial type: consistent, neutral, and misleading) ANOVA on average confidence ratings associated with final test performance. We found a main effect for trial type $F(2, 238) = 30.55$, p ,0.001, $\eta_p^2 = 0.20$, and an interaction between testing group and trial type, $F(2, 238) = 3.30$, $p < .05$, $\eta_p^2 = 0.03$. As Table 3 demonstrates, average confidence on consistent trials was higher when participants took an interpolated test ($M = 64.01$) then when they had not ($M = 54.83$), $t(121) = 2.62$, $p < .005$. This pattern was also demonstrated for misleading trials, where average confidence was higher when participants took an interpolated test ($M = 59.39$) than when they had not ($M = 49.45$), $t(121) = 2.55$, $p < .01$. Testing had no impact on average confidence associated with neutral trials. Additionally, warning had no impact on average confidence.

Experiment 2 discussion

We found that the warning did not improve accuracy on misleading trials, it did not reduce production of misleading details, and it had no impact on average confidence ratings. These data suggest that under these conditions where delays occur between the original event and post-event information, and between the post-event information and final test, participants may be less able to access original event details even when encouraged to do so. These results are discussed in the General Discussion.

General discussion

Two important findings emerged from the present study. We found that pre-retrieval warnings were effective in reducing RES, even in situations where the interpolated test was less helpful in strengthening original event memory. However, in Experiment 2, we found that the benefits of pre-retrieval warnings were eliminated when the warning was given immediately before the final test and 24 h after the post-event information. Further, the results from Experiment 1 suggest that while the interpolated test may not have supported long-term retention of originally learned information, it did potentiate the learning of the post-event information. That is, participants who were unwarned and had received an interpolated test, were more likely to produce misinformation on the final test of memory when compared to participants who did not take an interpolated test.

Prior research suggests that the enhanced learning of PEI resulting from testing may foster a bias at retrieval that is measured by enhanced production of misinformation (e.g[5]). Importantly, when participants were warned, they were less likely to report those suggested details, indicating that a warning may operate to counteract the bias that results from retrieval fluency. Additionally, the pre-retrieval warning improved memory accuracy for original event details in the interpolated testing group, suggesting that while original details may have not been better learned because of the delayed interpolated test (e.g[16]), they did remain accessible and could be reported when participants were informed about the quality of PEI.

In Experiment 2, when a 24-hour retention interval was inserted after the PEI and before the warning, the benefit of warnings in reducing RES was eliminated. That is, even when warned, participants who took an interpolated test were less accurate on the final test and more likely to produce misinformation on questions associated with misleading trials. Chan et al. (2022) argued that one reason why delayed warnings may be less effective is that, rather than operating to reduce biased retrieval and encourage more effective source monitoring (e.g[5]), warnings after the presentation of PEI may encourage participants to ignore or forget the PEI. A large body of literature on directed forgetting (DF) has shown that people can intentionally forget some previously learned information (for review see[20]).

The present study was not designed to test a directed forgetting account for pre-retrieval warnings. Rather, we sought to examine whether warnings could be effective when conditions of original memory consolidation, reactivation, learning of PEI, and reconsolidation were met. In fact, in such conditions as those established in Experiment 2, interpolated testing may have had a different kind of impact on original event information. With a delayed initial test, original event memory was given sufficient time to consolidate prior to reactivation. Researchers have shown that memories consolidate within about six hours of initial encoding[21,22]. Reactivation of original event memories through testing places those memories into a malleable state, increasing the likelihood of alteration by subsequent post-event information (cf[6]). By introducing a delay after the PEI and prior to the final test, we also allowed time for reconsolidation of those new memories. Under such conditions, warnings may not have been effective because the original event details were altered or no longer accessible. Future work may consider both directed forgetting and reconsolidation effects as factors that may constrain the efficacy of pre-retrieval warnings, as one of the limitations of the present work is that we do not disentangle these two possibilities.

Conclusions

Warnings may be a useful tool to improve eyewitness memory accuracy. As our research demonstrates, a warning that follows misleading PEI can be effective even in situations where learning of PEI may be enhanced due to witness questioning. Additionally, warnings may also improve over-confidence in memory reporting by encouraging witnesses to consider all sources of information that may influence their memories. However, pre-retrieval warnings may not be a general panacea for misinformation. As the present study demonstrates whether those warnings occur in proximity to present misinformation may determine its effectiveness. As additional research emerges, we will be able to develop general guidelines for just how warnings may be used to improve real world eyewitness reports.

Data availability

This research was conducted as part of Alia Wulff's dissertation and was supported by theNational Science Foundation Award Number:1728764; Award Title: Using Stress and Warningto Improve Eyewitness Memory. All data are available https://osf.io/3kf84/?view_only=adc91d71360249a09fcebc56b5ce0438.

Received: 18 May 2024; Accepted: 20 December 2024
Published online: 06 January 2025

References

1. Belli, R. F., Lindsay, D. S., Gales, M. S. & McCarthy, T. T. Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. *Mem. Cogn.* **22**(1), 40–54. https://doi.org/10.3758/BF03202760 (1994).
2. Loftus, E. F., Miller, D. G. & Burns, H. J. Semantic integration of verbal information into a visual memory. *J. Exp. Psychol. Hum. Learn. Mem.* **4**(1), 19–31. https://doi.org/10.1037/0278-7393.4.1.19 (1978).

3. Ost, J. et al. False memory ≠ false memory: DRM errors are unrelated to the misinformation effect. *PLoS ONE* **8**(4), e57939. https://doi.org/10.1371/journal.pone.0057939 (2013).
4. Chan, J. C. K., Thomas, A. K. & Bulevich, J. B. Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science* **20**(1), 66–73 (2009).
5. Thomas, A. K., Bulevich, J. B. & Chan, J. C. K. Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *J. Mem. Lang.* **63**(2), 149–157. https://doi.org/10.1016/j.jml.2010.04.004 (2010).
6. Chan, J. C. K. & LaPaglia, J. A. Impairing existing declarative memory in humans by disrupting reconsolidation. *Proc. Natl. Acad. Sci.* **110**(23), 9309–9313. https://doi.org/10.1073/pnas.1218472110 (2013).
7. Chan, J. C., Wilford, M. M. & Hughes, K. L. Retrieval can increase or decrease suggestibility depending on how memory is tested: The importance of source complexity. *J. Mem. Lang.* **67**(1), 78–85 (2012).
8. Chan, J. C. K., Manley, K. D. & Lang, K. Retrieval-enhanced suggestibility: A retrospective and a new investigation. *J. Appl. Res. Mem. Cogn.* **6**(3), 213–229. https://doi.org/10.1016/j.jarmac.2017.07.003 (2017).
9. Rindal, E. J., DeFranco, R. M., Rich, P. R. & Zaragoza, M. S. Does reactivating a witnessed memory increase its susceptibility to impairment by subsequent misinformation?. *J. Exp. Psychol. Learn. Mem. Cogn.* **42**(10), 1544–1558. https://doi.org/10.1037/xlm0000265 (2016).
10. Gordon, L. T. & Thomas, A. K. Testing potentiates new learning in the misinformation paradigm. *Mem. Cogn.* **42**(2), 186–197. https://doi.org/10.3758/s13421-013-0361-2 (2014).
11. Gordon, L. T. & Thomas, A. K. The forward effects of testing on eyewitness memory: The tension between suggestibility and learning. *J. Mem. Lang.* **95**, 190–199. https://doi.org/10.1016/j.jml.2017.04.004 (2017).
12. Thomas, A. K., Gordon, L. T., Cernasov, P. M. & Bulevich, J. B. The effect of testing can increase or decrease misinformation susceptibility depending on the retention interval. *Cogn. Res.: Princ. Implic.* **2**(1), 45. https://doi.org/10.1186/s41235-017-0081-4 (2017).
13. Blank, H. & Launay, C. How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *J. Appl. Res. Mem. Cogn.* **3**(2), 77–88. https://doi.org/10.1037/h0101798 (2014).
14. Karanian, J. M. et al. Protecting memory from misinformation: Warnings modulate cortical reinstatement during memory retrieval. *Proc. Natl. Acad. Sci. USA* **117**(37), 22771–22779. https://doi.org/10.1073/pnas.2008595117 (2020).
15. Chan, J. C. K., O'Donnell, R. & Manley, K. D. Warning weakens retrieval-enhanced suggestibility only when it is given shortly after misinformation: The critical importance of timing. *J. Exp. Psychol. Appl.* **28**(4), 694–716. https://doi.org/10.1037/xap0000394 (2022).
16. Pansky, A. Inoculation against forgetting: Advantages of immediate versus delayed initial testing due to superior verbatim accessibility. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**(6), 1792–1800. https://doi.org/10.1037/a0028460 (2012).
17. Chan, J. C. K. & Langley, M. M. Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *J. Exp. Psychol. Learn. Mem. Cogn.* **37**(1), 248–255. https://doi.org/10.1037/a0021204 (2011).
18. Bonham, A. J. & González-Vallejo, C. Assessment of calibration for reconstructed eye-witness memories. *Acta Psychol. (Amst)* **131**(1), 34–52. https://doi.org/10.1016/j.actpsy.2009.02.008 (2009).
19. Higham, P. A., Blank, H. & Luna, K. Effects of postwarning specificity on memory performance and confidence in the eyewitness misinformation paradigm. *J. Exp. Psychol. Appl.* **23**(4), 417–432. https://doi.org/10.1037/xap0000140 (2017).
20. Sahakyan, L. & Smith, J. R. 'A long time ago, in a context far, far away': Retrospective time estimates and internal context change. *J. Exp. Psychol. Learn. Mem. Cogn.* **40**(1), 86–93. https://doi.org/10.1037/a0034250 (2014).
21. Schiller, D. et al. Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature* **463**(7277), 49–53. https://doi.org/10.1038/nature08637 (2010).
22. Walker, M. P., Brakefield, T., Allan Hobson, J. & Stickgold, R. Dissociable stages of human memory consolidation and reconsolidation. *Nature* **425**(6958), 616–620. https://doi.org/10.1038/nature01930 (2003).

Acknowledgements
This project was supported by the National Science Foundation Award Number: 1728764; Award Title: Using Stress and Warning to Improve Eyewitness Memory.

Author contributions
A.W. collected the two reported experiments as part of their dissertation. A.W. developed the experiments, collected and analyzed data, and wrote initial drafts of the manuscript. J.K. was instrumental in conceptual, and theoretical interpretation, and reviewing data and drafts of the manuscripts. E.R. was instrumental in conceptual, and theoretical interpretation, and reviewing data and drafts of the manuscripts.A.T. was the primary supervisor on this dissertation research, the PI of the grant that funded this research, and wrote the draft of this submitted manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.K.T.

Reprints and permissions information is available at www.nature.com/reprints.

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