Memory & Cognition
https://doi.org/10.3758/s13421-024-01650-4

Timing of testing affects earwitness memory

McKinzey G. Torrance[1] · **Ayanna K. Thomas**[1]

Accepted: 2 October 2024

Abstract

The present experiments investigated how interpolated testing and postevent misinformation affected earwitness memory. We examined how the number of tests and when tests occurred affected memory for an overheard event and source monitoring. Across three experiments, participants overheard a crime (i.e., the witnessed event), heard a news report summarizing the crime (postevent information), took a cued-recall test, and lastly, took a source-monitoring test. Experiment 1 compared three groups: repeated cued-recall test, repeated listen, single cued-recall test. Participants in the interpolated test group received a test after the witnessed event and again after the postevent information. Participants in the repeated listen group heard the witnessed event, and immediately relistened to the event before being presented with the postevent information. In Experiment 2 and 3, we varied the retention intervals between the witnessed event and the postevent information in a repeated test context. Our findings suggest that when participants took a test before presentation of the postevent information, they were less accurate on a final cued-recall test of the witnessed event. Importantly, the timing of the first test in relation to the witnessed event and postevent information differentially affected memory for the witnessed event and source monitoring of event and postevent details.

Keywords Source monitoring · Misinformation · Cued-Recall · Event memory · Earwitness

Research has consistently demonstrated that postevent misinformation negatively affects eyewitness memory accuracy (Loftus, 2005). However, there is comparatively little research examining earwitness event memory. In one study comparing eyewitness to earwitness memory, when tested after a 4-day delay, earwitnesses had significantly poorer memory for a conversation compared with eyewitnesses (Campos & Alonso-Quecuty, 2006). The authors suggested that audio-only memories may be less rich in detail compared with audio-visual memories, which translated into less contextual information in memory to aid with retrieval. Critically, Campos and Alonso-Quecuty (2006) did not include any misleading postevent information in their study. Thus, it remains unclear how earwitness accuracy would be affected by the presentation of misleading postevent information.

The present study examined earwitness event/item and source memory in a misinformation paradigm where participants heard an audio-only witnessed event, were presented with postevent information in the form of an audio-only newscast, took a final memory test for the original event, and finally completed a source-monitoring test. Prior research suggests that source monitoring may be impaired when the two sources share the same modality. Our primary goal was to examine this possibility within the earwitness methodology (Campos & Alonso-Quecuty, 2006; Johnson, 2006). Importantly, the postevent information included details that were inconsistent with the original event.

Across experiments, we manipulated the number and timing of memory tests. We included an interpolated memory test that followed the witnessed event and preceded any postevent information to examine how the test influenced processing of postevent information and impacted later source monitoring. In an *eye*witness experimental design, researchers have demonstrated that an interpolated test sometimes increases susceptibility to postevent information resulting in less accuracy on final cued-recall or recognition tests as compared with when an interpolated test is not taken (for review, see Chan et al., 2017). That is, when participants were asked to retrieve witnessed event details, they were less likely to correctly retrieve those details if they had taken an interpolated test. Additionally, they were more likely to

✉ McKinzey G. Torrance
mckinzey.torrance@tufts.edu

[1] Tufts University, 490 Boston Ave, Medford, MA 02155, USA

retrieve postevent misinformation. However, research also suggests that interpolated testing may help to differentiate the two sources of information depending on the timing of that test in relation to the original and postevent information (Gordon & Thomas, 2017). The goal of our three experiments was to assess event memory and source monitoring for event and postevent details in an earwitness design where both the original and postevent information were heard.

Repeatedly testing witnesses

In some situations, interpolated testing in eyewitness memory experiments can promote eyewitness accuracy. Specifically, self-implemented interviews (Gabbert et al., 2012), immediate verbal recall of a face (LaPaglia & Chan, 2012), and verbatim memory recall (Pansky & Tenenboim, 2011) were found to improve eyewitness memory of the crime. These results may in part emerge because the interpolated test may improve learning of the witnessed event (backward effect of testing). However, there is also a large body of research that demonstrates that inserting a test between the witnessed event and postevent information results in reduced accuracy on a final test of the original event (Chan et al., 2012, 2017; Thomas et al., 2010, 2017). This is called retrieval enhanced suggestibility (RES), and research suggests that interpolated testing may influence how postevent details are learned and remembered (Gordon & Thomas, 2014, 2017; Thomas et al., 2010, 2017).

Unlike studies that have found benefits of interpolated testing, RES studies typically employ the same test following the witnessed event and after the postevent information. The interpolated test typically is a cued-recall or four-alternative forced-choice test. Following the interpolated test, postevent information is presented often as a retelling of the crime that includes novel misleading information that was not present during the crime. Research suggests that the interpolated test preceding postevent information may impact the learning of postevent details (forward effect of testing). For example, in one study, Gordon and Thomas (2014) found that when participants were asked to recall as many details as they could remember on a final cued-recall test following the postevent information, participants who took an interpolated test (e.g., a test before the postevent information) were more likely to recall both the original and postevent details as compared with participants who took one final test. These data suggest that interpolated testing, or testing that precedes the postevent narrative, may impact learning of that information.

Although postevent details may be better learned after interpolated testing, they may also be incorrectly retrieved as original event details on a final test of memory if participants are influenced by the ease with which those details come to mind (e.g., Benjamin et al., 1998; Thomas et al. 2010). Thomas and colleagues (2010) demonstrated that postevent details were retrieved more quickly than original details, and that difference was even greater when participants had taken an interpolated test.

Importantly, several studies have found that by promoting source monitoring, participants can reduce RES and respond correctly with witnessed event information. For example, in studies that warned participants that the postevent information may not be reliable, RES was significantly reduced or eliminated (Karanian et al., 2020; Thomas et al., 2010; Wulff, 2019, 2022). Researchers suggested that warnings may promote more effortful memory retrieval by encouraging source monitoring instead of relying on retrieval fluency to answer questions. Indeed, more effortful memory retrieval through source monitoring has been found to improve eyewitness memory (Lindsay & Johnson, 1989). In the present experiments, we measured memory for original event details and source memory associated with retrieved details to determine whether timing of interpolated tests would differentially impact performance on these two measures.

Remembering the source of auditory information

Using perceptual and contextual cues that differentiate one source from another improves source monitoring. For example, voice characteristics may be used to differentiate two audio sources. Indeed, across several studies, participants were able to identify the source of words that were read to them based on the gender of the speaker (Dodson et al., 1998; Dodson & Shimamura, 2000; Ferguson et al., 1992; Geiselman & Bellezza, 1976; Johnson et al., 1995; Kausler & Puckett, 1981; Lindsay et al., 1991). These studies concluded that gender of the voice was a memory characteristic people could reliably use to distinguish source (Dodson et al., 1998). The present study utilized different gendered voices in the witnessed event and postevent information to support differences in memory characteristics between the two events.

An interpolated test may also serve as a contextual element that could benefit source monitoring. Using the RES methodology Gordon and Thomas (2014) found that participants were more likely to direct attention to details in a postevent narrative if they had taken an interpolated test. Additionally, Gordon and Thomas (2017) found that participants who had taken an interpolated test were better able to identify discrepancies between the witnessed event and postevent information. These experiments suggest that when asked to consider ways in which the two sources of information differ, participants who took an interpolated test may be better equipped to assess differences. Importantly, few studies have examined source-monitoring accuracy when an interpolated test is designed to promote both backward effects of testing (i.e., Gabbert et al., 2012) and forward effects of testing (i.e., Gordon & Thomas, 2014). The present study examines this question withing the context of an earwitness paradigm.

The present study

The first experiment compared three groups: repeated cued-recall test, repeated listen, and single cued-recall test, to test the hypothesis that interpolated testing influenced processing of postevent details. The repeated listen group is a valuable comparison as it provides the opportunity to improve learning of the original event without impacting learning of postevent details. We predicted that participants who took an interpolated test would be more likely to retrieve postevent details on the final cued-recall test. However, when asked to monitor the source of retrieved details we expected that participants who took an interpolated test would demonstrate better ability to accurately remember source as compared with the other two groups. That is, we predict that the goal of each type of test will determine witness memory accuracy.

On the surface, predictions regarding the final memory test for details and the final source-monitoring test may seem in conflict. However, prior research suggests that a memory test that does not require participants to explicitly identify the source of retrieved details may be biased by retrieval fluency (Thomas et al., 2010). As interpolated testing has been shown to increase retrieval fluency of postevent details, we expected that retrieval fluency may bias responding on the final cued-recall test. However, when participants are explicitly asked to attribute a source to event details, we expect that they will consider perceptual and contextual details to support source monitoring decisions. Critically, lower accuracy on the final memory test but higher accuracy on the source-monitoring test would reflect that information from the original event is still accessible to participants, even if they had lower final memory-test accuracy compared with the other two groups.

The second experiment tested whether a 24-hour retention interval between the witnessed event and the interpolated memory test would influence the test-impacted processing of the postevent information. In line with our hypotheses from Experiment 1, we expected interpolated testing would result in lower final-test accuracy and higher misinformation production, as well as higher source-monitoring accuracy compared with the single cued-recall test group. These results would suggest that the forward effect of testing will emerge even when backward effects of interpolated testing may be reduced by the retention interval.

The third experiment varied the timing of a 24-hour retention interval to explore event memory and source monitoring when the interpolated test was delayed (e.g., Experiment 2) or when the final cued-recall test was delayed. The witnessed event and interpolated test were separated by a 24-hour retention interval, or they were presented in the same session with a retention interval between the postevent information and final memory tests. We compared these two groups to assess whether source monitoring would be disrupted when the witnessed event, interpolated test, and postevent information occurred in a single session that was delayed from final testing.

Experiment 1

Method

Participants

A priori power analyses were completed to determine sample size using G*Power (Faul et al., 2007). The planned statistical analysis which required the highest sample size to be appropriately powered (interaction for the 3 × 2 analysis of variance [ANOVA]) given a medium effect size (r = .25, η_p^2 = 0.04), determined the sample size (N = 158). Given the novel nature of this experiment, a medium effect size was chosen as a practical effect size to determine sample size. Furthermore, we conducted a sensitivity analysis after data collection to establish the true minimum effect sizes for the experiment and to assess if the analyses were appropriately powered.[1] Of note, the a priori power analysis was only calculated for the interaction of the 3 × 2 ANOVA and not for any of the other statistical tests.

A total of 220 participants were sampled from Prolific, an online site where vetted subjects can take part in virtual experiments. Participants were excluded from the experiment if they had not paid sufficient attention to the study. We checked attention through interpolated test accuracy, accuracy on two attention-check questions on the interpolated memory test and final memory test, and the duration they spent on the experiment as well as on specific webpages that contained the witnessed event or postevent information. Sixty-two participants were removed from analysis for not paying attention to the witnessed event, postevent information, or for having 0% accuracy on the interpolated memory test. Participants failed preestablished attention checks when they either advanced past the witnessed event before

[1] Using G*Power, we calculated minimum effect sizes at a power of .8 for the between-participants factor of group (η_p^2 = 0.033), the within-participants factor of detail type (η_p^2 = 0.011), and the within–between interaction of group and detail type (η_p^2 = 0.013). We calculated this effect size with a sample size of 150, given that G*Power assumes equal size groups and our lower sample size per group was 50. Our results had effect sizes that were larger than the minimum detectable effect size, which were all small-to-medium effects.

the recording was over (less than 390 s) or spent more time on the witnessed event webpage than needed to realize the recording was over and navigate to the next page (more than 500 s). Further if they spent less than 80 s on the postevent information webpage or more than 200 s, they were also excluded. A total of 158 participants were included in the following analysis. While participants were evenly distributed across groups during sampling, due to the exclusion criteria, there was an uneven number of participants for each group in the sample used for analysis. Specifically, there were 56 participants in the single test group, 52 were in the repeated listen group, and 50 were in the interpolated test group. All participants were between the ages of 18 and 65 years ($M = 37$, $SD = 12.64$). Eighty-six participants were under the age of 35, while 52 participants were 35 to 55 years old, and 20 participants who were older than 55 but younger than 65. Most of the sample (61%) identified as women, then 34% identified as men, and lastly 5% identified as nonbinary. Seventy percent of participants identified as White, 9% identified as East Asian, 8% identified as Black, 6% identified as having multiple racial identities, 4% identified as Latino and/or Hispanic, 2% identified as South Asian, and 1% identified as Southeast Asian. All participants were compensated at a rate of $8 per hour.

Materials

Witnessed event (conversation) The witnessed event is an approximately 6-minute audio clip of two men having a conversation. One of the men has a distinctly higher pitched voice than the other. The conversation details a robbery that one of the men committed earlier that day. During the audio, there are two female voices that present noncritical information (e.g., the participant enters the bar and sits down for a drink). Participants read the following instructions prior to listening to the event: "You will now listen to an audio narrative of a conversation. Please make sure your computer volume is at a level where you can clearly hear what is being played. Do not start the audio narrative unless you are in a distraction-free area or have headphones on. The audio narrative will take around 6 minutes to complete. Only advance to the next section of the survey once the audio narrative is finished. Once you are ready to start the audio narrative, please click the video player below."

Interpolated cued-recall test The interpolated cued-recall test was a forced-response cued-recall test. Participants were instructed, "You will now answer questions about the conversation you heard earlier. Please answer to the best of your abilities while avoiding responding you don't know or don't remember. Please only write one answer for each question." The test consisted of ten questions about specific details from the witnessed event. One example of a question is, "What mask was one of the speakers wearing during the robbery?" Each question was presented in chronological order of when the related detail was mentioned in the witnessed event. Participants typed their answer for each question. The first question and last question in the test were used as attention checks for the experiment and were not included in the final analysis. Each question in the memory test related to details from the audio-only witnessed event. These details were experimentally manipulated in the postevent information to either be neutral and not confirm what was said in the audio-only witnessed event or be misleading and have a changed detail from what was heard in the witnessed event (Appendix 1).

Visual search task The visual search task consisted of three hidden picture images (see https://www.printablee.com). This task was one of the filler tasks completed during the experiment. Each of the images was shown for 2 minutes. Participants were instructed, "You will be shown 3 photos with hidden items in them. The list of hidden items is provided at the bottom of each of the photos. You will be given 2 minutes to study each of the photos and find the hidden items. After you study one of the photos and find the hidden items, you will be automatically forwarded to the next screen to report how many of the items you found. Then, you will be shown the next photo." After 2 minutes, participants were advanced to a second screen where they typed in how many objects they found in the picture.

Sudoku The filler task was two games of Sudoku for participants to complete at their own pace. There was a timer counting down 10 minutes at the top of the web page. Participants would be automatically moved to the next section of the experiment after the 10 minutes were up. Even if participants finished the Sudoku puzzles early, they would not be moved to the next section until the timer had reached zero. The sudoku puzzles were one of the filler tasks completed during the experiment.

Postevent information (news report) The postevent information was an approximately 1.5 minute retelling of the witnessed event. The information was styled as a radio news report presented by a female reporter. The postevent information is told by only one female voice. The report did not present information in the same order as how participants heard it in the witnessed event. The postevent information contained eight critical details. These critical details were related to the questions asked in the interpolated memory test and what participants were asked again on the final memory test. Half of the details were neutral and neither confirmed nor contradicted what participants had heard in the witnessed event (i.e., "The assailants stole a large sum of money."). Half of the details had been changed to something different than what participants heard in the witnessed event (i.e., "The assailants stole $10,000."). There were two versions of this postevent information created such that the misleading and neutral details were counterbalanced across the two versions (Appendix 2, Appendix 3). One of the two versions was randomly assigned to participants prior to listening.

Final cued-recall test The final cued-recall test was a forced cued-recall test identical to the interpolated memory test. Before the test, the participants read, "You will now answer some questions about *the conversation* you heard at the beginning of the experiment. Please only write one answer for each question and answer every question. Answer to the best of your abilities while avoiding responding you don't know or don't remember." As with the interpolated test, each question on the final cued-recall test related to details from the witnessed event, which had also been mentioned in the postevent information. Thus, all questions on the final memory test were about items from the witnessed event that had been mentioned in the interpolated memory test or the postevent information. As described above, the critical details were manipulated to either be neutral or misleading.

Source-monitoring test The source-monitoring test was a separate test from the final cued-recall test. Participants were given the instructions, "You will now see your answers for each question on the memory test. You will be asked to distinguish where you remembered each answer from: The original conversation, the news report, both sources, or neither." Participants were presented with each of their answers from the final cued-recall test. Participants then indicated where they remembered that answer from, the witnessed event, postevent information, both sources, or neither source. Of note, this design does not allow us to fully study the source-monitoring process due to source attributions only being made for information participants produced themselves. Instead, this design was specifically chosen to assess the impact of interpolated testing and retention intervals on the source attributions of produced information. A source-monitoring test based on participant answers allowed us to measure source monitoring accuracy for information participants produced themselves instead of experimenter items from the witnessed event which participants may not have encoded. Further, our design allowed for the analysis of what types of misattributions (i.e., attributing misinformation to the witnessed event, correct information from the postevent information, or information coming from both sources) were occurring.

Scoring

Both the interpolated cued-recall test and the final cued-recall test were coded by two research assistants. The research assistants who scored the test varied across the three experiments. The assistants were given a list of correct responses and scored questions correct if the answer was on that list. Assistants scored responses of "I don't know" or "I don't remember" as incorrect. The same two research assistants that scored correct answers also coded for misleading details, neutral details, and "I don't know" responses produced on the interpolated and final memory tests. There were no discrepancies between the two raters' responses on either of the memory tests.

When evaluating the source of "I don't know" responses or guesses on the final memory test, a source-monitoring decision was scored as correct if participants labeled these responses a coming from "neither" the crime video nor the news report. Analyses on the production of "I don't know" or "I don't remember" responses as well as misleading and neutral detail production can be found in the Supplemental Materials (Tables 3–5).

Procedure

From the online Prolific posting, participants were directed to a Qualtrics survey. All participants confirmed they qualified for the experiment and signed the informed consent before being given directions to plug in headphones and adjust their volume to a comfortable listening level. Participants read that the audio would be approximately 6-minute long and that they should not advance to the next page before the audio was finished. There were no instructions indicating that participants would be tested on the content of the audio narrative later. When participants were ready, they played the witnessed event. The participants listened to the witnessed event once, before advancing to the next screen where they listened to the witnessed event a second time. Participants listened to the witnessed event twice to support encoding of the initial event. Multiple exposures to support encoding is a common practice for stimuli that may be less well remembered (Cabeza & St Jacques, 2007; Ebbinghaus, 1964). This second screen had the same instructions as the first. Following the witnessed event, all participants completed the 6-minute visual search task. Participants were then randomly assigned to one of three groups (Fig. 1). In the single test group, participants played sudoku for 6 minutes. In the repeated listen group, participants listened to the witnessed event for a third time. In the interpolated test group, participants took the interpolated memory test. Once participants completed the interpolated test, they could advance to the next section. After participants finished their respective tasks, all participants had 10 minutes to play two sudoku games. Participants in the single test group had 10 minutes added to their original 6 minutes. Participants were automatically advanced to the next screen after 10 minutes. The following screen instructed participants to listen to a news report about

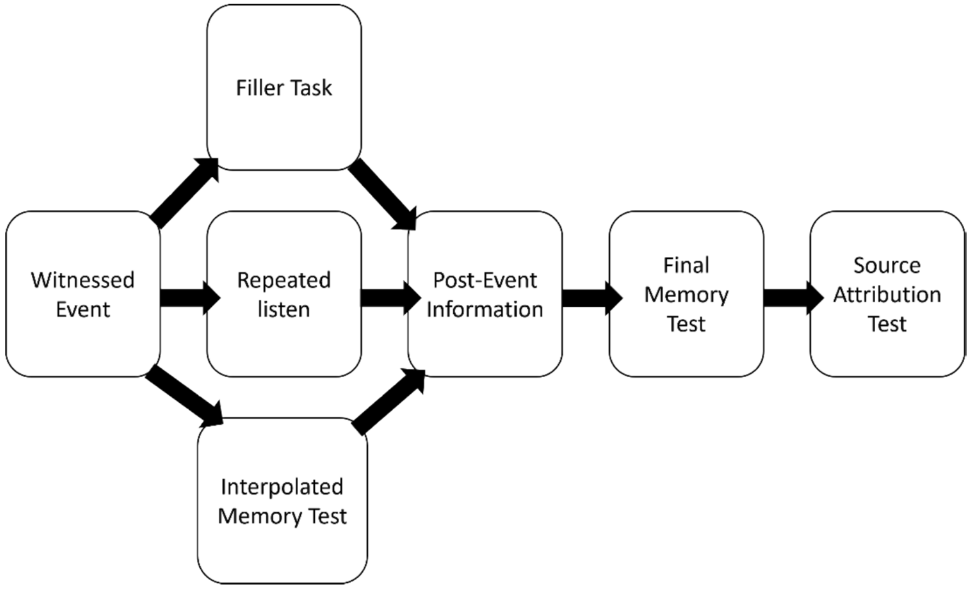

Fig. 1 Experiment 1 procedure

the crime they had overheard. Participants started the news report when they were ready. After listening to the postevent information, all participants took the final memory test, which was identical to the interpolated memory test. Finally, participants completed the source-monitoring test. They finished the experiment by answering demographic questions about their age, gender, race, level of education, when they learned English and number of psychology classes they have taken in their lifetime. They were debriefed about the experiment and were given our contact information in case they had any questions or concerns. Lastly, they received a code for compensation through the Prolific website.

Results

All post hoc comparisons used Bonferroni correction as indicated by notation (*). Interpolated and final test accuracy by experiment are included in the Supplemental Materials (Tables 1 and 2). Of note, initial memory-test accuracy and final memory-test accuracy did not seem to greatly differ across age groups (see Supplemental Materials, Tables 6–11). We also examined the *z* scores for initial memory-test and final memory-test accuracy and found no outliers that may be unduly manipulating results, relating to participant age or otherwise. These findings were consistent across all experiments.

Interpolated cued-recall test accuracy

We ran a *t* test looking at the difference in interpolated test accuracy by detail type (neutral, misleading). Detail type refers to the questions in the interpolated cued-recall test that could be answered with details from the crime as well as details from the postevent information. Importantly, in the postevent information, the details were changed to either be neutral (neither confirm nor contradict the contents of the witnessed event) or misleading (contradict the findings of the witnessed event). This analysis was conducted to check that the misinformation presented in the postevent information was not naturally occurring as answers prior to the presentation of the postevent information. Accuracy on interpolated test questions relating to misleading details (M = 0.75, SD = 0.31) was not significantly different from question accuracy associated with neutral details (M = 0.75, SD = 0.29), $t(98)$ = 0.08, p = .933, d = 0.02. Average accuracy on the interpolated test across detail type was approximately 75% (SD = 0.29).

Final cued-recall test accuracy

We ran a 3 × 2 ANOVA looking at the effects of group (single test, repeated listen, interpolated test), detail type (misleading, neutral), and their interaction, on final test accuracy. Detail type affected final test accuracy, $F(1,155) = 24.76$, $p < .001$; $\eta_p^2 = 0.14$. Accuracy on misleading questions (M = 0.56, SD = 0.49) was significantly lower compared with accuracy on neutral questions ($M = 0.69$, $SD = 0.46$), $t(157) = -4.96$, $p < .001$*, $d = -0.44$. There was no main effect of group $F(2,155) = 0.17$, $p = .838$, $\eta_p^2 < .001$, nor significant interaction between group and detail type, $F(2,155) = 1.68$, $p = .190$, $\eta_p^2 = 0.02$.

Final cued-recall test—Misinformation production

We ran a one-way ANOVA looking at the effect of group (single test, repeated listen, interpolated test) on misinformation production on final memory test. Misinformation production is when participants produce misinformation that they encountered during the postevent information and report it as an answer on the final memory test for the witnessed event. Importantly, it is not just any incorrect information participants produced on the final memory test, but specifically misinformation they had encountered in the postevent informant. Group significantly affected misinformation production, $F(2,155) = 3.67$, $p = .03$; $\eta_p^2 = 0.05$ (Fig. 2). Participants in the single test group produced significantly less misinformation ($M = 0.11$, $SD = 0.32$) compared with participants in the interpolated test group ($M = 0.19$, $SD = 0.39$), $t(155) = -2.54$; $p = .04$*, $d = -0.47$. There was no significant difference between the single test group and the repeated listen group, $t(155) = -0.38$, $p = 1.000$*, $d = -0.08$, nor the repeated listen and the interpolated test groups, $t(155) = -2.12$, $p = .106$*, $d = -0.41$.

Source monitoring

We calculated a Fisher exact test to examine if participant group (single test, repeated listen, interpolated test) influenced which source attribution (witnessed event, postevent information, neither, both) for misinformation was produced on the final memory test. A Fisher exact test was conducted because the Fisher exact test is not sensitive to small expected values. Analysis of source attributions of correct responses can be found in the Supplemental Materials (Table 12).

There was a significant association between group and source attributions for misinformation produced on the final memory test ($p = .04$; Table 1). Participants who received an interpolated test attributed the misinformation produced on the final test to the postevent information more (62/74) than statistically expected and attributed postevent information to the witnessed event less (0/74) than statistically expected. The participants in the single test group attributed produced misinformation to the postevent information less (34/50) than statistically expected and attributed postevent information to the witnessed event more (5/50) than statistically expected. The participants in the repeated listen group attributed misinformation to the postevent information less (34/51) than statistically expected and attributed postevent information to the witnessed event more (4/51) than statistically expected.

Experiment 1 discussion

Experiment 1 was a novel exploration of the effect of repeated testing on earwitness event memory and source monitoring of event and postevent information. There was an effect of misinformation exposure, such that all participants had significantly lower accuracy for misleading questions compared with neutral questions on the final cued-recall test. Additionally, participants who took an interpolated test produced significantly more

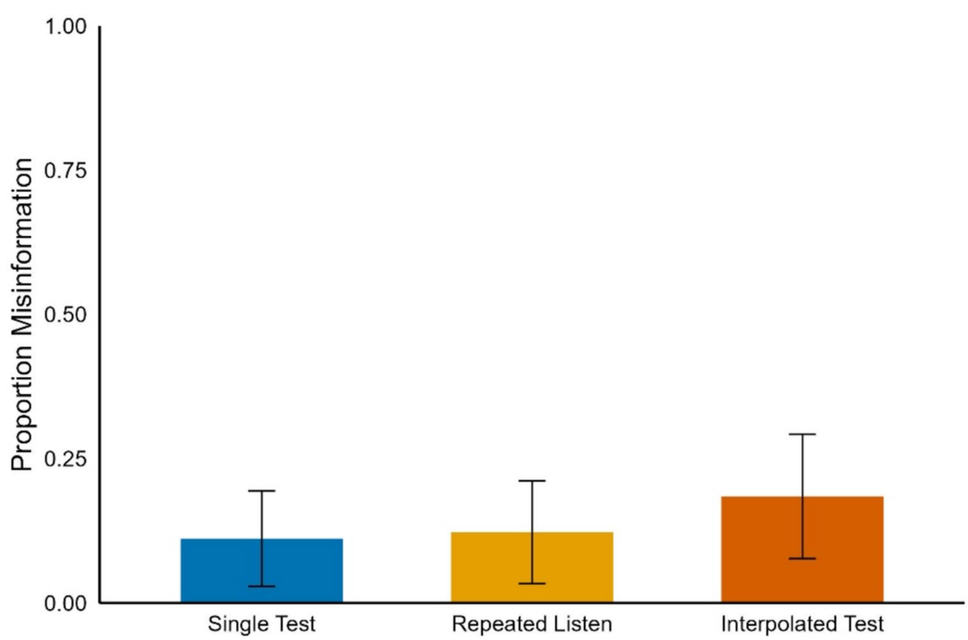

Fig. 2 Experiment 1 misinformation production on the final memory test by group. *Note.* Bars represent means and lines represent the 95% confidence interval

Table 1 Experiment 1 source monitoring responses by group

Group		Source				Total
		Conversation	News report	Neither	Both	
Single test	Observed count	5	34	2	9	50
	Expected count	2.57	37.14	1.14	9.14	50
Repeated listen	Observed count	4	34	1	12	51
	Expected count	2.62	37.89	1.17	9.32	51
Interpolated test	Observed count	0	62	1	11	74
	Expected count	3.81	54.97	1.69	13.53	74
Total	Observed count	9	130	4	32	185
	Expected count	9	130	4	32	185

The above data reflect source monitoring responses on the four questions of the final memory test where misinformation could be produced. The observed counts represent the collected data. Expected counts represent the distribution of counts given the total and marginal totals observed from the collected data as if there were no relationship between group and source

misinformation on the final test compared with participants in the single test or repeated listen groups, extending RES to audio only events (Chan et al., 2009, 2017; Karanian et al., 2020; Thomas et al., 2010).

Group affected source attributions. Participants who took an interpolated test were more likely to identify misleading postevent information as being from the news report. Furthermore, they attributed none of the misleading postevent information they produced on the final memory test to the witnessed event. In contrast, participants in the single test and repeated listening conditions reported more of the misleading postevent information they produced on the final memory test as being from the witnessed event, and less as being from the news report. Consistent with the forward effect of testing, these results suggest that when participants took an interpolated memory test, that they better learned the subsequently presented information from the postevent information. On the final memory test participants responded with this better learned information leading to higher rates of misinformation production compared with the other two conditions. Yet during a source-monitoring task, which prompted participants to consider the source of their responses, participants in the interpolated test condition were able to attribute that misinformation to the correct source (i.e., the news report).

Of note, these results also provide context as to why there were no differences between the groups on memory accuracy for misleading questions. Specifically, the memory-accuracy and source-monitoring results suggest that participants who received an interpolated test produced more misinformation on the final memory test, while participants in the other two condition produced more noncritical incorrect information.

Experiment 2

Experiment 2 explored the role of interpolated testing after a 24-hour retention interval between the witnessed event and postevent information. The repeated listen group was not included in Experiment 2, to focus solely on the role of interpolated testing in earwitness item and source memory. We introduced a retention interval between the witnessed event and postevent information for two reasons. First, by placing the interpolated test immediately before the postevent information we were able to capitalize on the influence the test should have on processing that information. Second, we predicted that the retention interval would help to further differentiate the two audio sources (i.e., witnessed event and postevent information). Past research found that when sources of information are made distinct from one another through a delay for 48 hours, that participants made more accurate source judgments compared with when sources were less temporally distinct from one another (Lindsay, 1990). Therefore, we predicted that participants would be better able to attribute their final test answers to the correct source as both the retention interval and interpolated test were combined to generate cues that would be diagnostic of the appropriate source.

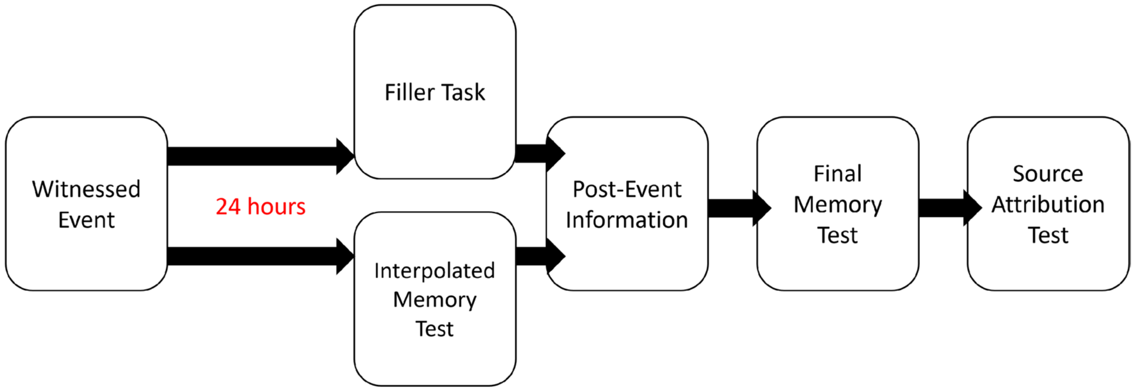

Fig. 3 Experiment 2 procedure

Methods

Participants

Based on the number of participants sampled in Experiment 1, roughly the same number of participants were sampled for Experiment 2. We completed a sensitivity analysis after data collection was finished to check the minimum detectable effect sizes.[2] We determined an effect-size sensitivity analysis would be more appropriate than a power analysis given the previous study was the first of its kind and the population effect size is unknown (Giner-Sorolla et al., 2019). Thus, an effect-size sensitivity analysis would calculate the minimum population effect size detectable at or above the desired power established in the first study (0.8). Two-hundred and twelve participants were sampled from Qualtrics for Part 1 of the experiment. In Part 1, 31 participants failed the same attention checks established in Experiment 1. These participants were not invited back for Part 2. One hundred and eighty-one participants were invited back for Part 2. There was an approximately 25% attrition rate. Of the remaining 139 participants, one was removed from analysis for 0% accuracy on the interpolated test, and seven participants were removed for reporting they purposely used the postevent information to answer final test questions instead of thinking back to the original witnessed event. One hundred and twenty-nine participants were included in the following analysis. There were 70 participants in the single test (filler task) group and 59 participants in the interpolated test group. All participants were between the ages of 18 and 65 years (M = 37, SD = 11.98). Sixty-one participants were under the age of 35, while 51 participants were between the ages of 35 and 55, and 12 participants were between the ages of 55 and 65. Fifty-nine percent of the participants identified as women, 36% identified as men, 4% identified as nonbinary, and approximately 1% identified as trans men. The majority of the sample identified as White (79%), 11% had multiple racial identities, 4% identified as East Asian, 2% identified as Black, 2% identified as Southeast Asian, and 1% identified as South Asian. All participants were compensated at a rate of $8 per hour.

[2] Using G*Power, we calculated an effect-size sensitivity analysis for the main effects and interaction of the planned 2 (group) × 2 (detail type) ANOVA on final test accuracy. We used a sample size of 118 given that the interpolated test group only have 59 participants and G*Power assumes equal sizes between groups. We calculated minimum effect sizes at a power of .8 for the between-participants factor of group ($\eta_p^2 = 0.033$), the within-participants factor of detail type ($\eta_p^2 = 0.013$), and the within–between interaction of group and detail type (η_p^2= 0.013). Our results had effect sizes that were larger than the minimum detectable effect size, which were all small-to-medium effects.

Materials

The same materials used in Experiment 1 were used in Experiment 2.

Procedure

From our online Prolific posting, participants were directed to a Qualtrics survey. Participants followed the same procedure as described in Experiment 1, except after listening to the witnessed event for a second time, participants had a 24-hour retention interval before returning to the Prolific website. Participants were invited by their Prolific ID to take part in the second session of the experiment. Upon return, participants in the single test group completed a 6-minute visual search task. In the interpolated test group, participants returned to Prolific and took the interpolated memory test (Fig. 3).

Results

All post hoc comparisons used Bonferroni correction as indicated by notation (*).

Interpolated cued-recall test accuracy

We ran a t test to examine the effect of detail type (neutral, misleading) on interpolated test accuracy. Accuracy on interpolated test questions relating to misleading trials ($M = .57$, $SD = 0.30$) was not significantly different from accuracy on neutral trials ($M = .57$, $SD = 0.28$), $t(116) = 0$, $p = 1.000$, $d = 0$. Accuracy on the interpolated test across detail type was approximately 57% ($SD = .29$).

Final memory test accuracy

We ran a 2 × 2 ANOVA looking at the effects of group (single test, interpolated test), detail type (misleading, neutral), and their interaction on final test accuracy. Detail type significantly affected final test accuracy, $F(1,127) = 48.05$, $p < .001$; $\eta_p^2 = 0.27$. There was significantly lower accuracy on questions associated with misleading details ($M = 0.40$, $SD = 0.49$) compared with those associated with neutral details ($M = 0.58$, $SD = 0.49$), $t(128) = -6.82$, $p < .001$*, $d = 0.61$. There was also a significant interaction between group and detail type, $F(1,127) = 5.09$, $p = .02$; $\eta_p^2 = 0.04$ (Fig. 4). Specifically, participants in the single test group had higher accuracy on questions associated with misleading details ($M = 0.57$, $SD = 0.29$) compared with participants in the interpolated test group ($M = 0.33$, $SD = 0.31$), $t(127) = 2.46$; $p = 0.02$*, $d = 0.43$. There were no significant differences between the groups for neutral questions, $t(127) = -0.29$, $p = .772$*, $d = 0.05$. There was no main effect of group, $F(1,127) = 2.67$, $p = .104$, $\eta_p^2 = 0.02$.

Final cued-recall test—Misinformation production

We ran an independent-samples t test looking at the effect of group (single test, interpolated test) on misinformation production on the final memory test. Participants in the single test group produced less misinformation ($M = 0.16$, $SD = 0.14$) compared with participants in the interpolated test group ($M = 0.26$, $SD = 0.16$), $t(127) = -3.61$, $p < .001$, $d = -0.64$ (Fig. 5).

Source monitoring

We calculated a Fisher exact test to examine if participant group (single test, interpolated test) was associated with source attribution (witnessed event, postevent information, neither, both) for misinformation produced on the final memory test. Analysis of source attributions of correct details can be found in the Supplemental Materials (Table 13).

There was a significant association between group and source attributions for misinformation produced on the final memory test ($p < .001$; Table 2). Participants who received an interpolated test attributed the misinformation produced on the final test to the postevent information more (103/122) than statistically expected and to the witnessed event less (4/122) than statistically expected. The participants in the single test group attributed their misinformation to the postevent information less (55/91) than statistically expected and to the witnessed event more (12/91) than statistically expected.

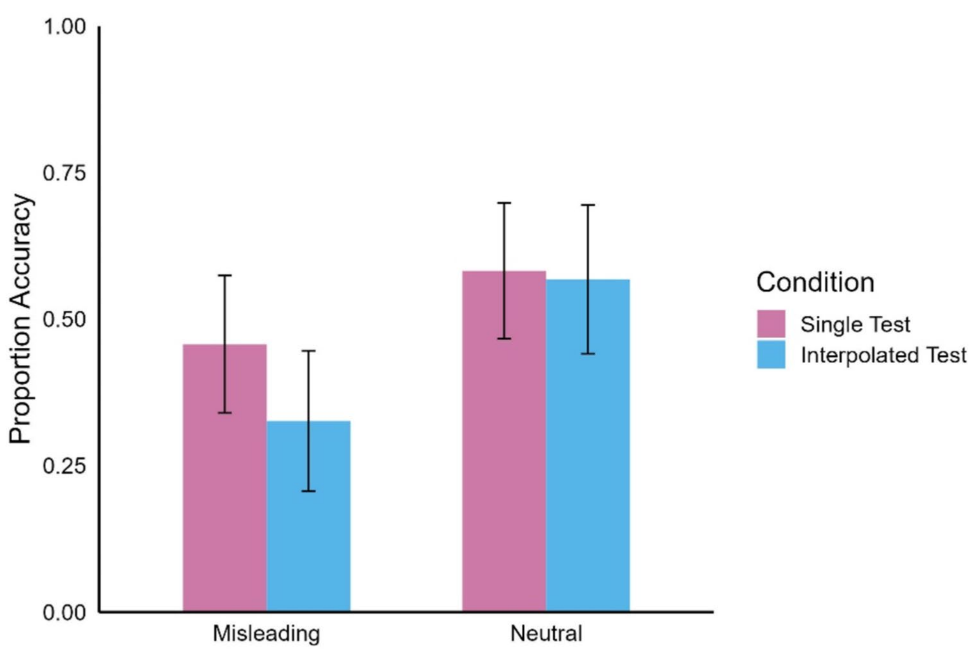

Fig. 4 Experiment 2 final test accuracy by group and detail type. *Note.* Bars represent means and lines represent 95% confidence interval

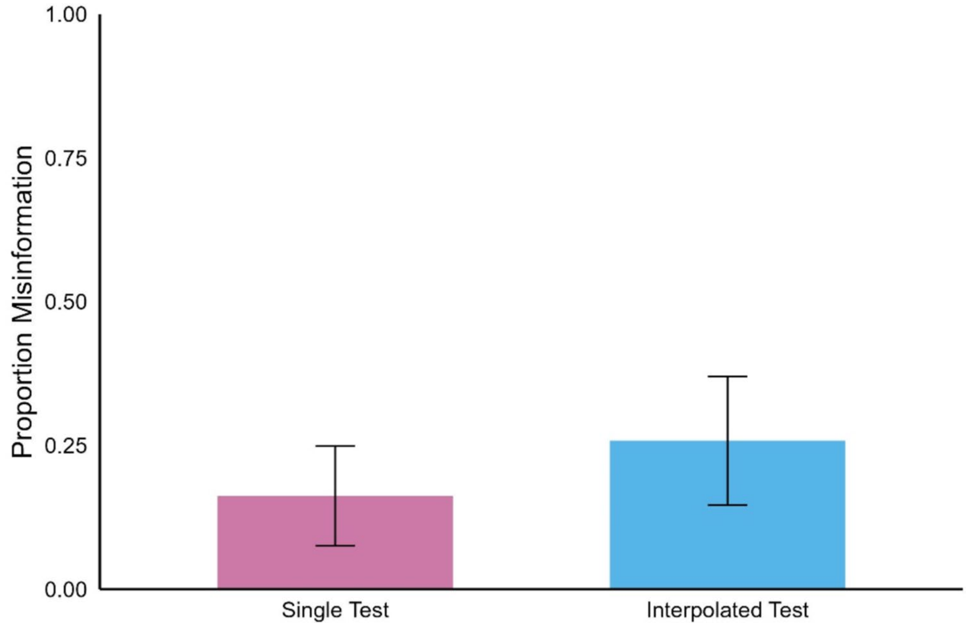

Fig. 5 Experiment 2 misinformation production on the final memory test by group. *Note.* Bars represent means and lines represent 95% confidence interval

Table 2 Experiment 2 source monitoring responses by group

Group		Source				
		Conversation	News report	Neither	Both	Total
Single test	Observed count	12	55	5	19	91
	Expected count	6.84	67.50	2.56	14.10	91
Interpolated test	Observed count	4	103	1	14	122
	Expected count	9.16	90.50	3.44	18.90	122
Total	Observed count	16	158	6	33	213
	Expected count	16	158	6	33	213

The above data reflect source monitoring responses on the four questions of the final memory test where misinformation could be produced. The observed counts represent the collected data. Expected counts represent the distribution of counts given the total and marginal totals observed from the collected data as if there were no relationship between group and source

Experiment 2 discussion

Experiment 2 explored the effect of interpolated testing on earwitness event memory and source monitoring when there was a 24-hour retention interval between the witnessed event and postevent information. Participants who took an interpolated test had significantly lower memory accuracy on questions relating to misleading details compared with participants in the single test condition. Consistent with Experiment 1, they also produced more misinformation on the final test compared with participants who did not take an interpolated test. Further, participants attributed more of the misinformation they produced on the final memory test to the postevent information, while participants in the single test group attributed less of the misinformation they produced to the postevent information and more to the witnessed event. These data support the conclusion that interpolated testing may result in biased responding on the final test that may be corrected when participants are asked to engage in source monitoring in an earwitness paradigm.

Experiment 3

Experiment 3 explored the effect of interpolated testing on event memory and source monitoring when the witnessed event and postevent information were separated by a 24-hour retention interval, as compared with when they occurred in the same session. In Experiments 1 and 2, participants who took the interpolated test had poorer performance on the final cued-recall test but were more likely to correctly attribute produced misinformation to the correct source. However, when the witnessed event, interpolated memory test, and postevent information are presented in a single session prior to a delay, participants could have difficulty distinguishing between the two sources of information. The close presentation in time between the two events may increase the similarity in memory characteristics between the two events, reducing source monitoring accuracy. Experiment 3 directly compares learning in a context that promotes the extraction of diagnostic cues (both temporal and perceptual) and in a context that limits extraction of these cues. We predicted that presenting the witnessed event and postevent information in a single session will result in lower memory accuracy and lower source monitoring accuracy compared with when the witnessed event and postevent information are temporally distinct.

Method

Participants

As with Experiment 2, roughly the same number of participants were sampled from Prolific. We completed a sensitivity analysis after data collection to check the minimum effect size detectable for the analyses.[3] Two hundred and fifteen participants were sampled from Prolific. Thirty-one participants were not invited back to the second session of the experiment for failing attention checks in Part 1 of the experiment. These were the same attention checks used in Experiments 1 and 2, and they were split between Parts 1 and 2 of the present experiment. One hundred and ninety-five participants were invited back for Part 2. There was an attrition rate of 23% for Session 2. One-hundred and fifty-one participants completed Session 2 of the experiment. Eight additional participants were removed from analysis for failing attention checks in Session 2 of the experiment. The following analyses were conducted on a sample of 143 participants. There were 69 participants in the Delay—Test 1 group and 74 participants in the Delay—Test 2 group. The age of participants ranged from 19 to 65 years (M = 35, SD = 11.72). Seventy-eight participants were under 35 years old, while 54 participants were 35–55 years old, and 11 participants were over 55 years old. On average, participants were 35 years old (SD = 11.72). Most participants identified as female (62%). Thirty-six percent identified as men, and 2% identified as nonbinary. Most participants also identified as White (65%). Ten percent identified as Black, 10% had multiple racial identities, 6% identified as Latino and/or Hispanic, 4% identified as East Asian, 3% identified as Southeast Asian, and 2% identified as South Asian. All participants were compensated at a rate of $8 per hour.

Materials

The same materials were used from Experiments 1 and 2.

Procedure

We compared two groups: the Delay—Test 1 group, and the Delay—Test 2 group (Fig. 6). The name of each group is a reference to which occurred first, a retention interval, and then what occurred after the delay, the interpolated memory test, *or* the final memory test. In the Delay—Test 1 group there is a 24-hour retention interval before the interpolated memory test. In the Delay—Test 2 group, the witnessed event, interpolated test, and postevent information preceded a 24-hour retention interval, which was followed by the final memory test.

Results

All post hoc comparisons used Bonferroni correction as indicated by notation (*).

Interpolated cued-recall test accuracy

We ran a 2 × 2 ANOVA looking at the effects of group (Delay—Test 1, Delay—Test 2), detail type (misleading, neutral), and their interaction on interpolated test accuracy. Group significantly affected interpolated test accuracy, $F(1,141) = 5.16$, $p = .03$, $\eta_p^2 = 0.04$. Specifically, participants in the Delay—Test 1 group had significantly lower interpolated test accuracy ($M = 0.64$, $SD = 0.48$) compared with participants in the Delay—Test 2 group ($M = 0.73$, $SD = 0.44$), $t(141) = -2.27$, $p = .03$*, $d = -.32$. This is expected given that the Delay—Test 1 group took the interpolated memory test after a 24-hour retention interval after the witnessed event, while the Delay—Test 2 group took the interpolated memory test in the same session as hearing the witnessed event. There was not a main effect for detail type, $F(141) = 1.90$, $p = .169$, $\eta_p^2 = 0.01$, nor a significant interaction, $F(141) = 0.01$, $p = .928$, $\eta_p^2 < 0.001$.

[3] Using G*Power, we calculated an effect-size sensitivity analysis for the main effects and interaction of the planned 2 (group) × 2 (detail type) ANOVA on final test accuracy. We used a sample size of 138 given that the Delay – Test 1 group had a sample of 69 participants and G*Power assumes equal sample sizes across groups. We calculated minimum effect sizes at a power of .8 for the between participants factor of group ($\eta_p^2 = 0.033$), the within participants factor of detail type ($\eta_p^2 = 0.011$), and the within-between interaction of group and detail type ($\eta_p^2 = 0.011$). Our results had effect sizes that were larger than the minimum detectable effect size, which were all small-to-medium effects.

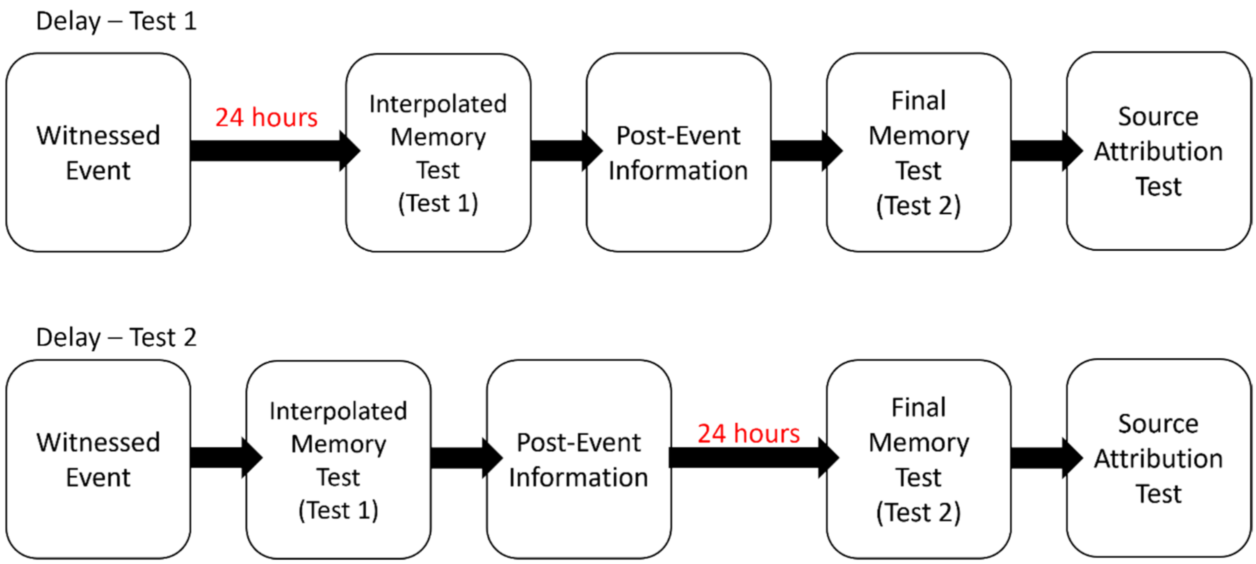

Fig. 6 Experiment 3 procedure

Final cued-recall test accuracy

We ran a 2 × 2 ANOVA looking at the effects of group (Delay—Test 1, Delay—Test 2), detail type (misleading, neutral), and their interaction on final test accuracy. Group significantly affected final test accuracy, $F(1,142) = 14.02$, $p < .001$, $\eta_p^2 = 0.09$. Specifically, participants in the Delay—Test 1 group had significantly lower final recall accuracy ($M = 0.49$, $SD = 0.50$) compared with participants in the Delay—Test 2 group ($M = 0.66$, $SD = 0.47$), $t(141) = -3.75$, $p < .001$*, $d = -0.63$. Detail type was found to significantly affect final test accuracy, $F(1,142) = 24.45$, $p < .001$, $\eta_p^2 = 0.15$. Participants had significantly lower accuracy on questions associated with misleading details ($M = 0.51$, $SD = 0.50$) compared with questions associated with neutral details ($M = 0.64$, $SD = 0.48$), $t(142) = -4.73$, $p < .001$*, $d = -0.38$. There was also a significant interaction between group and detail type, $F(1,142) = 13.32$, $p < .001$, $\eta_p^2 = 0.09$

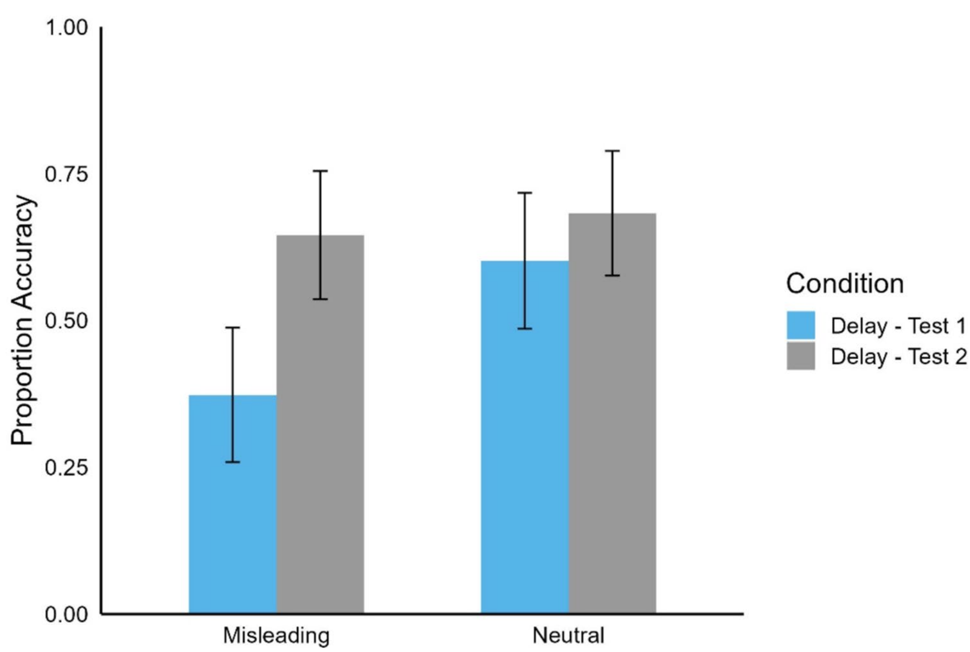

Fig. 7 Experiment 3 final test accuracy by group and detail type. *Note.* Bars represent means and lines represent 95% confidence intervals

(Fig. 7). Participants in the Delay—Test 1 group had lower accuracy on questions associated with misleading details ($M = 0.37$, $SD = 0.33$) compared with participants in the Delay—Test 2 group ($M = 0.64$, $SD = 0.32$), $t(141) = -5.04$, $p < .001$*, $d = -0.84$. There were no significant differences between the groups for neutral questions, $t(141) = -1.50$, $p = .134$*, $d = -0.25$.

Final cued-recall test—Misinformation production

We ran an independent-samples t test looking at the effect of group (Delay—Test 1; Delay—Test 2) on misinformation production on the final memory test. Participants in the Delay—Test 1 group produced more misinformation ($M = 0.22$, $SD = 0.15$) as compared with participants in the Delay—Test 2 group ($M = 0.11$, $SD = 0.13$), $t(141) = 4.21$, $p < 0.001$, $d = 0.70$ (Fig. 8).

Source monitoring

We calculated a Fisher exact test to examine if participants group (Delay—Test 1, Delay—Test 2) was associated with source attribution (witnessed event, postevent information, neither, both) for misinformation produced on the final memory test. Analysis of the source attributions for correct details on the final memory test can be found in the Supplemental Materials (Table 14).

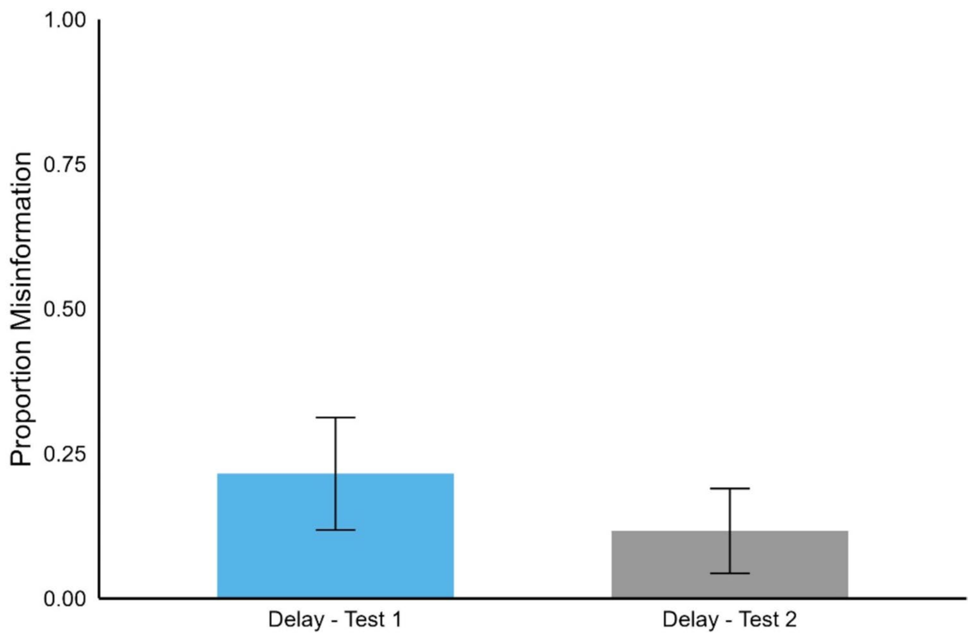

Fig. 8 Experiment 3 misinformation production on the final memory test by group. *Note.* Bars represent means and lines represent 95% confidence intervals

Table 3 Experiment 3 source monitoring responses by group

Group		Source				
		Conversation	News report	Neither	Both	Total
Delay-Test 1	Observed count	4	104	2	9	119
	Expected count	14.56	93.05	1.90	9.49	119
Delay-Test 2	Observed count	19	43	1	6	69
	Expected count	8.44	53.95	1.10	5.51	69
Total	Observed count	23	147	3	15	188
	Expected count	23	147	3	15	188

The above data reflect source monitoring responses on the four questions of the final memory test where misinformation could be produced. The observed counts represent the collected data. Expected counts represent the distribution of counts given the total and marginal totals observed from the collected data as if there were no relationship between group and source

There was a significant association between group and source attributions for misinformation produced on the final memory test (p < .001; Table 3). Specifically, participants in the Delay—Test 1 group attributed *more* of the misinformation they produced (104/119) to the postevent information than statistically expected. These data replicate the findings in Experiment 2. The Delay—Test 2 group, on the other hand, attributed *less* of the misinformation they produced (43/69) to the postevent information than statistically expected. Further, participants in the Delay—Test 1 group attributed *less* of the misinformation they produced (4/119) to the witnessed event than statistically expected, while the Delay—Test 2 group attributed *more* of their misinformation produced (19/69) than statistically expected.

Experiment 3 discussion

When a retention interval followed the witnessed event, event memory accuracy was negatively affected as compared with when a retention interval followed the postevent information. In fact, accuracy was not significantly different on misleading and neutral trials when the final test was delayed. However, even though event memory performance was better, participants who were presented with the witnessed event and postevent information in the same session were more likely to misattribute retrieved postevent details to the witnessed event than statistically expected and demonstrated greater source misattribution errors than participants who had a delay after the witnessed event. When the witnessed event, interpolated test, and postevent information occurred in the same session the test was not effective in promoting diagnostic source monitoring cues.

General discussion

The present experiments investigated how engaging in memory retrieval impacted earwitness event memory and source monitoring in an earwitness misinformation paradigm. We found that the RES effect replicated in the earwitness paradigm. Repeated memory retrieval encouraged the learning of event details, and in some cases the source of those details, depending on when testing occurred.

When people took an interpolated memory test, they were more likely to produce misleading details on a final test of memory than when they did not take an interpolated test. In Experiment 1, participants in the interpolated test group seemed to monitor the source of the misinformation produced on the final memory test more accurately compared with the single test and repeated listen groups. Paired together, these results suggest that interpolated testing increased the learning of postevent details and the source of those details. This finding is consistent with results from the eyewitness memory literature that found that while an interpolated memory test may result in RES that this impact is significantly reduced when participants are prompted to consider the source of information (Karanian et al., 2020).

In Experiment 2, when a retention interval was inserted between the witnessed event and interpolated test, participants who received an interpolated test were more likely to produce misinformation on the final cued-recall test than participants who had not taken the interpolated test. Additionally, these participants were more likely to correctly attribute the misinformation they produced on the final cued-recall test to the postevent information than participants who did not take an interpolated test. Importantly, results from Experiment 2 demonstrated that the retention interval alone did not promote effective source monitoring of remembered misleading details. That is, participants in the single cued-recall test group were more likely to make source attribution errors of remembered misleading details as compared with participants in the interpolated test group. It is the interpolated test which seems to support learning of contextual information needed to distinguish misinformation as coming from the news report.

Consistent with Experiment 2, participants in Experiment 3 that had a delayed interpolated test (Delay—Test 1) had high rates of misinformation production on the final cued-recall test accompanied by an increased ability to correctly attribute misinformation to the postevent information. Interestingly, for participants who had a delayed final memory test (Delay—Test 2), RES was eliminated. That is, there was no difference in cued-recall accuracy between neutral and misleading trials. As demonstrated in prior research (Thomas et al., 2017), when the witnessed event, interpolated test, and postevent information are presented in the session before the delay, participants may be less likely to rely on retrieval fluency associated with postevent details. When the fluency bias is reduced, participants may be more likely to demonstrate the benefits of the interpolated test.

Although participants demonstrated similar levels of accuracy on neutral and misleading trials in the Delay—Test 2 group, they did still produce some misinformation on the cued-recall test. Interestingly, for these produced misleading details, participants *incorrectly* attributed those details to the original event. That is, participants who had a delayed final test demonstrated better item memory but poorer source memory than participants who had a delay between the witnessed event and interpolated test. These results suggest that postevent details may have been incorporated into the witnessed event memory, and therefore, participants incorrectly attributed that misinformation to the witnessed event. Alternatively, the delay after postevent information may have resulted in more forgetting of source specifying cues. Research suggests that source characteristics are more quickly forgotten compared with event details (Braun, 1999;

Brown & Halliday, 1991; Schacter et al., 1984) and this differential forgetting may be exacerbated for audio events (Campos & Alonso-Quecuty, 2006). Thus, participants may have higher memory accuracy on the final cued-recall test but misattribute misinformation to the witnessed event because source details were being more quickly forgotten.

Another explanation of these results is that it may be difficult to extract diagnostic cues from audio-only sources that occur in the same testing session. Past voice identification research suggests that gender of speaker is a reliable marker for source monitoring. While the present study's sources (i.e., the witnessed event and postevent information) differed on several source characteristics (e.g., gender of speaker), participants may have difficulty remembering or effectively using these audio characteristics at longer retention intervals.

Conclusion

Past research has focused on eyewitness memory accuracy and reliability resulting in a gap in our understanding of earwitness memory capabilities. The present study found that interpolated testing conferred several advantages to earwitness memory in the presence of misinformation. Consistent with past eyewitness work, interpolated testing resulted in better memory for the original event and an improved ability to differentiate between the original event and postevent information. However, these advantages were dependent on when the interpolated test occurred in the sequence of events, and whether participants were encouraged to engage in source monitoring.

Appendix 1

Critical details associated with questions on the final cued-recall test

Details from the crime	Misinformation from the news report
Nixon mask	Clinton mask
Rikers Island	San Quinton
Russian mob	Local Mafia
Face	Arm
Old security guard	Store manager
10–15 minutes	5 minutes
Punched him	Kicked him
$30,000	$10,000

Appendix 2

Postevent information (Counterbalance A)

Today a store was robbed of <u>10,000 dollars</u>. Even though police responded to the robbery <u>quickly</u>, the assailants got away. Only one injury has been reported from the event. Thankfully, <u>the store manager</u> is recovering from being pistol whipped by one of the robbers at the local hospital. Two men have been apprehended in connection with the robbery. They were found leaving a local restaurant together. One of the men is believed to have been at the robbery and was disguised <u>in a Mask</u>. The police are currently looking for the other suspects involved with the robbery. The other man is believed to be an organizer of the crime. Through questioning these two individuals the police have obtained some information about the other four assailants that are still on the run. One of these suspects is believed to have served time in <u>Prison</u>. Further, another one of the suspects may be working with <u>the local Mafia</u> in the area. Finally, another suspect is known to **have a scar**. Street cameras caught this suspect being <u>kicked</u> by a fellow assailant directly following the robbery. While the cause of the infighting is unknown, police suspect the suspect to be visibly hurt. Please contact the police if you have any additional information regarding this crime or the assailants which could help lead to their capture. Thank you.

Appendix 3

Postevent Information (Counterbalance B)

Today a store was robbed of <u>a large sum of money</u>. Even though police responded to the robbery <u>in 5 minutes</u>, the assailants got away. Only one injury has been reported from the event. Thankfully, <u>the victim</u> is recovering from being pistol whipped by one of the robbers at the local hospital. Two men have been apprehended in connection with the robbery. They were found leaving a local restaurant together. One of the men was believed to have been at the robbery and was disguised in a <u>Clinton mask</u>. The other man is believed to be an organizer of the crime. Through questioning these two individuals the police have obtained some information about the other four assailants that are still on the run. One of these suspects is believed to have served time in <u>San Quentin</u>. Further, another one of the suspects may be working with <u>the local crime syndicate</u> in the area. One suspect is known to have <u>an arm scar</u>. Street cameras caught this suspect being <u>punished</u>

by a fellow assailant directly following the robbery. While the cause of the infighting is unknown, police suspect the suspect to be visibly hurt. Please contact the police if you have any additional information regarding this crime or the assailants which could help lead to their capture. Thank you.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13421-024-01650-4.

Author contributions M.T.: Conceptualization, methodology, formal analysis, data curation, visualization, funding acquisition, writing—original draft, reviewing, and editing.

A.T.: Conceptualization, methodology, supervision, writing—reviewing and editing.

Funding This work was supported by the International Honor Society in Psychology's Research Grant Contest Spring 2022; the Tufts University Graduate Student Research Competition Spring 2021.

Availability of data and materials None of the experiments in the present study was preregistered. All data, materials, and supplemental materials are available at the following repository: https://osf.io/pa6gd/?view_only=482649c358534f6490801516d745b711

Code availability Analysis code is available at: https://osf.io/pa6gd/?view_only=482649c358534f6490801516d745b711

Declarations

Ethics approval The present study was approved by the Tufts University Social, Behavioral & Educational Research Institutional Review Board.

Consent to participants All participants gave their informed consent prior to participation in the present study.

Consent for publication Not applicable.

Conflicts of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

References

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 14.

Braun, K. A. (1999). Postexperience advertising effects on consumer memory. *Journal of Consumer Research, 25*(4), 319–334. https://doi.org/10.1086/209542

Brown, A. S., & Halliday, H. E. (1991). Crytomnesia and source memory difficulties. *The American Journal of Psychology, 104*(4), 475–490. https://doi.org/10.2307/1422937

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences, 11*(5), 219–227. https://doi.org/10.1016/j.tics.2007.02.005

Campos, L., & Alonso-Quecuty, M. (2006). Remembering a criminal conversation: Beyond eyewitness testimony. *Memory, 14*(1), 27–36. https://doi.org/10.1080/09658210444000476

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science, 20*(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., Wilford, M. M., & Hughes, K. L. (2012). Retrieval can increase or decrease suggestibility depending on how memory is tested: The importance of source complexity. *Journal of Memory and Language*, *67*(1), 78–85. https://doi.org/10.1016/j.jml.2012.02.006

Chrobak, Q. M., & Zaragoza, M. S. (2012). The misinformation effect: Past research and recent advances. In A. M. Ridley, F. Gabbert, & D. J. La Rooy (Eds.), *Suggestibility in legal contexts* (pp. 21–44). John Wiley & Sons Ltd. https://doi.org/10.1002/9781118432907.ch2

Dodson, C. S., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1121–1136. https://doi.org/10.1037/0278-7393.24.5.1121

Dodson, C. S., & Shimamura, A. P. (2000). Differential effects of cue dependency on item and source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 1023–1044. https://doi.org/10.1037/0278-7393.26.4.1023

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology (H. Ruger & C. Bussenius, Trans.)*. Dover.

Faul, F., Erdfelder, E., Lang, A.-G., & Axel, B. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age differences in using source-relevant cues. *Psychology and Aging, 7*(3), 443–452. https://doi.org/10.1037/0882-7974.7.3.443

Gabbert, H. L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading postevent information with a self-administered interview. *Applied Cognitive Psychology, 26*, 568–575.

Geiselman, R. E., & Bellezza, F. S. (1976). Long-term memory for speaker's voice and source location. *Memory & Cognition, 4*(5), 483–489. https://doi.org/10.3758/BF03213208

Giner-Sorolla, R., Carpenter, T., Lewis, N. A., Montoya, A. K., Aberson, C. L., Bostyn, D. H., ..., & Schoemann, A. M. (2019). *Power to predict what? Considerations for planning and evaluating sample size.* SPSP Power Analysis Working Group.

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

Johnson, M. K. (2006). Memory and reality. *American Psychologist, 61*(8), 760–771. https://doi.org/10.1037/0003-066X.61.8.760

Johnson, M. K., De Leonardis, D. M., Hashtroudi, S., & Ferguson, S. A. (1995). Aging and single versus multiple cues in source monitoring. *Psychology and Aging, 10*(4), 507–517. https://doi.org/10.1037/0882-7974.10.4.507

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, 117*(37), 22771–22779. https://doi.org/10.1073/pnas.2008595117

Kausler, D. H., & Puckett, J. M. (1981). Adult age differences in memory for sex of voice. *Journal of Gerontology, 36*(1), 44–50. https://doi.org/10.1093/geronj/36.1.44

LaPaglia, J. A., & Chan, J. C. K. (2012). Retrieval does not always enhance suggestibility: Testing can improve witness identification performance. *Law and Human Behavior, 36*(6), 478–487. https://doi.org/10.1037/h0093931

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6), 7.

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

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*(3), 297–318. https://doi.org/10.1016/0022-0965(91)90065-Z

Loftus, E. F. (2005). Elizabeth F. Loftus. In G. Lindzey & W. M. Runyan (Eds.), *A history of psychology in autobiography* (vol. 9, pp. 199–227). American Psychological Association. https://doi.org/10.1037/11571-006

Loftus, E. F., Burns, H. J., & Miller, D. G. (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

Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing. *Memory & Cognition, 39*(1), 155–170. https://doi.org/10.3758/s13421-010-0005-8

Schacter, D. L., Harbluk, J. L., & McLachlan, D. R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior, 23*(5), 593–611. https://doi.org/10.1016/S0022-5371(84)90373-6

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., 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*(1), 45. https://doi.org/10.1186/s41235-017-0081-4

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

Wulff, A. N. (2019). *Warning placement affects eyewitness accuracy and confidence in the presence of misinformation [Master's thesis]*. Tufts University.

Wulff, A. N. (2022). *Repeated retrieval impacts memory for both original event and postevent information memory [Doctoral dissertation]*. Tufts University.