Understanding the Effects of Misinformation on Eyewitness Memory in Older Adults

A thesis submitted by

Himanshu Chaudhary

in partial fulfillment of requirements for the degree of

Master of Science

in

Psychology

Tufts University

February 2026

Advisors:

Elizabeth Race

Ayanna K. Thomas

Abstract

Eyewitness memory accuracy is influenced by both memory and metamemorial processes. Two factors that affect these cognitive processes and the reliability of eyewitness memory are age and exposure to misleading post-event information (PEI). The current study examined how providing metamemorial support through warnings and the opportunity to withhold answers impacted older adults' memory and metamemory performance in the face of misinformation. In two experiments, older adults were presented with a video depicting a non-violent crime and were subsequently presented with an auditory summary of that video (post-event information; PEI) containing misleading details. Participants were randomly assigned to either receive a warning after the PEI or to not receive a warning. Participant's performance on a forced report and free report memory test, as well as confidence estimates corresponding to their answers, were evaluated for memory and metamemory performance. I tested the impact of warning across two experiments using different versions of the forced report memory test. Experiment 1 employed a 4 alternative force choice (4-AFC) memory test, that increased the likelihood of familiarity-based responding. Experiment 2 employed a cued-recall test that required more engagement of recollective processes and effortful retrieval. Results demonstrated that older adults exhibited the misinformation effect across both experiments. However, older adults were able to effectively control their memory output by withholding incorrect and/or misleading details. Warnings did not improve older adults' memory and metamemorial processes.

Acknowledgements

I would like to extend my sincerest appreciation to my amazing advisors Dr. Elizabeth Race and Dr. Ayanna K. Thomas. Without their constant support, mentorship, and guidance, this would not have been possible.

I would also like to thank Dr. John Bulevich for their feedback and coordination as a committee member.

Additionally, I would like to thank all the members of my two labs: the Metacognition and Applied Memory Lab and the Integrative Cognitive Neuroscience lab for their friendship, support and feedback throughout this process.

Finally, thank you to my family, friends, and my partner, Ishita, for their love and support

Table of Contents

List of Figures

Understanding the Effects of Misinformation on Eyewitness Memory in Older Adults

Eyewitness testimony can be fallible and can have real-world consequences (Greene & Loftus, 1984). For decades, research has shown that eyewitness memory is prone to distortion and errors (Loftus, 1981) and is vulnerable to the exposure of misleading post-event information (PEI) that impairs the memory of the original event (Loftus et al., 1978). Incorrect PEI has been well-established to interfere with event memory, a distortion referred to as the misinformation effect (for review see, Loftus, 2005; McCloskey & Zaragoza, 1985). This vulnerability is sometimes exacerbated by age, with past work showing that older adults are especially susceptible to memory errors caused by misinformation (Bulevich & Thomas, 2012; Mitchell et al., 2003; Roediger & Geraci, 2007; Bornstein et al., 1995; Mueller-Johnson & Ceci, 2004). Two methods prevalent in prior work that have been shown to be effective in reducing susceptibility to misinformation are warnings (for review see, Blank & Launay, 2014) and the ability to volunteer or withhold retrieved information (Bulevich & Thomas. 2012). However, the two have not been utilized together in order to examine how they influence memory retrieval and impact one another to influence metamemorial processes in older adults.

Retention, Monitoring & Control

In the courtroom, eyewitnesses on the stand are asked to tell "the truth, the whole truth, and nothing but the truth". To do so, a witness must consider a multitude of things such as evaluating what they remember, how confident they are in their memory, and many others before deciding to either answer the question or refrain from doing so. Koriat and Goldsmith (1996) presented a framework that allowed for examining the independent contributions of these factors to overall memory performance. Within the Quality-Accuracy Performance (QAP) framework, Koriat and Goldsmith allowed for the measurement of three primary factors: retention,

metamemorial monitoring, and metamemorial control. Retention is the ability to retrieve previously learned information (i.e., recalling what details one had witnessed from a crime). Once details have been recalled, the next step is monitoring, which is the process of evaluating the retrieved information (i.e., confidence). The final step is control, which is the strategic regulation of memory output, or the strategic decision making of whether to volunteer or withhold responses. This framework allows for an empirical investigation of the experience of retrieving and reporting information under various external incentives.

Older adults have been shown to have similar monitoring abilities compared to younger adults when task demands are low (Daniels et al., 2009). Daniels et al. (2009) demonstrated this by having younger and older adults study words, make judgments of learning (JOLs), and later complete memory tests (recognition or cued recall). Specifically, in less demanding recognition tasks, older adults showed similar accuracy in their JOLs compared to younger adults. However, in the more cognitively demanding cued-recall tasks, which required participants to generate items from partial cues, older adults exhibited reduced monitoring accuracy. This finding suggests that while older adults can effectively monitor their learning under simple retrieval conditions, increased cognitive demands associated with tasks like cued recall significantly diminish their monitoring accuracy (Daniels et al., 2009).

Older adults have also been shown to demonstrate reliable metacognitive control when provided scaffolding for their memory. For example, older adults have been shown to strategically regulate their memory by selectively allocating cognitive resources toward high-value information, often through deliberate and controlled processes that compensate for declines in cognitive capacity associated with aging (Castel, 2008; Murphy & Castel, 2024). Specifically, research indicates that older adults become adept at managing memory deficits by prioritizing

retrieval processes that target highly valuable or meaningful details at the expense of less critical information when given a task to do so (Murphy & Castel, 2024). This selective memory approach is supported by metacognitive judgments and value-directed remembering paradigms, where older adults demonstrate the capability to strategically control their attention and retrieval efforts in alignment with task demands and the perceived value of information (Murphy & Castel, 2024; Castel, 2008).

These studies demonstrate that older adults have intact monitoring and intact control. These processes are crucial in an eyewitness memory task where participants are required to accurately remember a prior event. In such a task, participants are asked to rate the confidence of their recollections and exercise control over their answers by withholding or submitting details. Participants may better withhold incorrect answers and submit correct answers if they are able to accurately rate their confidence in those answers as low. In fact, research has consistently demonstrated that monitoring affects control. For example, Recent work by DeCaro and Thomas (2020) demonstrated that when older adults were explicitly prompted to retrieve target information prior to making judgments of learning (JOLs), the retrieval attempts influenced both their reaction times and the likelihood of selecting items for further study. Specifically, participants spent more time attempting retrieval for items they judged as difficult (low JOLs), and these same items were more frequently selected for restudy. This indicates a direct relationship between monitoring (JOLs) and subsequent control decisions (restudy choices), suggesting that the retrieval attempts prompted by monitoring judgments directly guide learners' strategic regulation of study resources.

Importantly, this relationship between monitoring and control decisions should generalize to eyewitness memory tasks, whereby confidence judgments guide older adults' decisions about

whether to withhold or report retrieved information. Indeed, Bulevich and Thomas (2012) demonstrated that older adults' decisions to withhold retrieved information in eyewitness tasks were directly mediated by their confidence levels, underscoring that confidence serves as a critical monitoring mechanism guiding strategic regulation during retrieval. Thus, despite differences in task measures, confidence judgments should similarly inform older adults' withholding decisions in the current study.

QAP and Misinformation in Older Adults

Koriat and Goldsmith's (1996) QAP framework allows researchers to directly examine the relationship between retention, monitoring and control. Memory retention is evaluated by measuring performance on the forced report test. Monitoring performance is measured by examining the relationship between the confidence and accuracy of participants' answers. Control performance is assessed by the ability to withhold incorrect answers on the free report test. This task additionally allows for the calculation of the level of confidence above which participants choose to submit an answer or submission criterion. Thus, researchers can observe how retention monitoring and control are impacted by misinformation through its differential impact on these measures.

Prior work utilizing the QAP framework with older adults demonstrated that the presence of misinformation or deceptive items resulted in a diminished ability to effectively monitor and control one's memory (Kelley & Sahakyan, 2001; Jacoby et al., 2005; Bulevich & Thomas, 2012). Additionally, Pansky et al. (2009) found that in a naturalistic memory task, older adults had lower overall memory performance than young adults. Older adults had a more liberal submission criterion and had reduced control sensitivity in weeding out errors. Researchers also found that older adults volunteered more incorrect details.

This pattern was further examined in the context of an eyewitness misinformation paradigm by Bulevich and Thomas (2012). In their first experiment, participants watched an episode of a TV show to serve as the witnessed event. Participants were then given a synopsis of the episode that they watched (PEI) which contained misleading and consistent details. Participants then took a forced report memory test, where they were required to answer every question. They also provided confidence judgements for each answer. Following participants took a free report test where they had the opportunity to withhold answers previously given on the forced report test.

Older adults displayed lower retention than younger adults, and both groups showed reduced retention on misleading items, as demonstrated by performance on the forced report test. In line with prior work (Kelley & Sahakyan, 2001; Dodson et al., 2007; Dodson & Krueger, 2006), Bulevich and Thomas (2012) found that older adults demonstrated disruptions in monitoring efficacy on misleading trials, demonstrating a pattern of overconfidence. Older adults were also less likely to withhold incorrect details, known as gains, on misleading trials, indicating a lack of effective control. Additionally, older participants' submission criteria were lower than those of younger adults, indicating that older adults were less selective in choosing answers to report, potentially increasing their risk of reporting inaccurate or misleading information. This is particularly important as it highlights how shifts in monitoring and control strategies with age may compromise the accuracy of eyewitness testimony, underscoring the need for further interventions to guide older adults' memory reporting decisions.

In Experiment 2, Bulevich and Thomas (2012) did just that by provided participants with detailed supportive instructions regarding how to make memory decisions. The "supportive instructions" contain detailed instructions on how to best recall details, with an emphasis on

context, as well as how to evaluate a memory's accuracy. The aim of these supportive instructions was to increase retrieval effort in order to encourage the effective use of source-specifying cues. Retrieval effort describes the cognitive work invested in bringing a memory to consciousness (Buckner et al., 1998; Barzykowski et al., 2021). It encompasses both the difficulty and duration of the retrieval attempt, with greater effort reflecting a deeper, more exhaustive search process. When supportive instructions were paired with the use of a cued recall task, older participants' memory and metamemorial performance on misleading trials drastically improved. Older adults no longer showed a mismatch of confidence and accuracy in misleading trials, they demonstrated gains, and their submission criteria changed to effectively withhold misleading information on the free report test. The results of this study's two experiments demonstrate that older adults' ability to monitor and control seems to be impaired as compared to younger adults by the presence of misinformation. However, older adults have the capacity to monitor and control effectively when given additional support to do so through supportive instructions and cued recall.

Using a General Warning to Resist Misinformation

Thomas et al. (2010) found that when participants were warned about the unreliability of the source of post-event information, younger participants were better able to resist misinformation. Similar warnings have been shown to be effective in reducing susceptibility to misinformation across various studies (for meta-analysis see, Blank & Launay, 2014). Karanian et al. (2020) investigated the mechanisms by which these warnings reduce susceptibility to misinformation using neuroimaging techniques and found evidence that warnings influence not only memory retrieval but also the mental reinstatement of an original event's context or source memory. In Karanian et al. (2020) participants first watched a clip from a film which served as

the original event. After an initial memory test, participants were placed in an fMRI scanner to listen to an audio narrative that summarized the clip and contained misleading details. This was done so that the original event and PEI would result in differential brain activations, with the original event engaging visual regions and the PEI engaging auditory regions. Crucially, some participants received a warning prior to the final memory test. By observing brain activity as participants answered questions, the study showed that the effective use of warning to reduce misinformation susceptibility was associated with the reactivation of brain regions associated with visual activity and the reduction of activation of brain regions associated with auditory activity. Simply, warned participants were more likely to activate brain regions associated with the accurate source more often than participants who were not warned. Warnings also have been shown to reduce confidence in memory on all trials, including misleading ones (Karanian et al., 2020). This could be because it made the participants more cautious and unsure of their answers.

Warnings differ from supportive instructions in their specificity and directiveness. Supportive instructions explicitly guide individuals toward effective retrieval strategies, such as emphasizing source-specifying cues and deeper memory searches, thus enhancing retrieval effort (Bulevich & Thomas, 2012). In contrast, warnings function as more general, indirect prompts intended to induce caution without providing explicit retrieval strategies. Despite this difference, I hypothesized that warnings could similarly influence older adults' metamemorial control and monitoring by increasing their wariness of misinformation. By serving as a situational demand, warnings may prompt individuals to more thoroughly evaluate retrieved details, thus raising their submission criterion. Consequently, I expected warnings to produce benefits analogous to supportive instructions by indirectly encouraging strategic regulation during memory reporting, particularly on misleading trials where older adults might otherwise be overly confident and less

selective. I thus aimed to evaluate the effect of warnings on retention, monitoring, and control in the current study.

Present Study

The present study utilized the QAP methodology paired with warnings in a standard misinformation experimental design to examine whether warnings can improve memory, monitoring performance, and control of memory in older adults. Older adult participants watched a short clip from a movie without dialogue in which a variety of non-violent crimes are depicted (the original event). After a short filler task, participants were asked to listen to an audio narrative. The audio narrative, which served as the post-event information, is a summary of the details from the original event and contains details that are either consistent with the original event, misleading (conflicting with the original event), or neutral (adding no additional information). Participants were randomly assigned to one of two warning groups prior to the task: post-warning, where the warning was presented after the post-event information, and no warning, where no warning was presented to the participant. After receiving all the necessary instructions for their assigned condition, the participants took a 4-AFC recognition memory test (Experiment 1) or a cued recall test (Experiment 2) about their memory for the original event. After each question participants were asked to rate their confidence in their answers. The forced report memory test was then followed by a free report test, where participants were re-presented with their answers from the forced report test and asked to submit only answers they were certain were correct.

I chose to utilize a 4-AFC recognition test for Experiment 1 because this format specifically targets familiarity-based biases in confidence, thereby allowing for the investigation of whether warnings can attenuate these biases. According to Yonelinas (2001), recognition

judgments can be based either on recollection, involving the conscious retrieval of specific episodic details, or on familiarity, a strength-based sense of prior exposure. Familiarity is particularly influential when recognition judgments are based on choosing among multiple provided alternatives, as familiar but incorrect lures artificially elevate confidence ratings (Yonelinas, 2001). Consequently, a 4-AFC recognition task challenges accurate metamemorial monitoring by introducing plausible and highly accessible misinformation options, inflating participants' confidence and thus impairing their ability to effectively evaluate their own memory accuracy. This biased monitoring can negatively impact subsequent control decisions, as inflated confidence makes participants more likely to submit incorrect details rather than withhold them. Therefore, implementing a 4-AFC recognition test in Experiment 1 allows me to assess whether warnings can specifically reduce these familiarity-based biases in confidence and, consequently, lead to improved monitoring and control in older adults.

In comparison, cued recall, as used in Experiment 2, still requires a response on every trial but induces greater retrieval effort. Unlike recognition, where participants often rely on the familiarity of options, cued recall requires the individual to actively generate a response, reducing the opportunity for familiarity-based responding. As a result, the differences in monitoring and control performance between forced and free report tests may be less pronounced with cued recall, since participants are already engaging in more effortful retrieval from the outset. Prior work (Bulevich & Thomas, 2012) has shown that when older adults are encouraged to exert more effort during retrieval, through the use of supportive instructions in addition to cued recall, they are better able to monitor and control their memory performance. By using a cued recall format in Experiment 2, I aimed to test whether warnings might support older adults' post-retrieval monitoring by encouraging them to more effortfully consider the source of their

memory. In this way, warnings may help reduce the likelihood that participants commit a source monitoring error, or misattribution of a suggested item from the PEI's source to the original event. If warnings prompt more cautious responses, I expect to see improvements not just in monitoring but also in control, with older adults setting a more conservative submission criterion and withholding misinformation more effectively.

Experiment 1

I hypothesized that warnings would be effective in improving retention on misleading trials on the forced report memory test by reducing misinformation susceptibility. Prior research (e.g., Thomas et al., 2010) has indicated that a warning provided prior to retrieval can reduce susceptibility to misinformation within a recognition task in younger adults. I hypothesized that older adults would be able to utilize these same warnings to show benefits on retention as younger adults did. I also hypothesized that warnings would improve monitoring, specifically the relationship between monitoring (confidence) and accuracy on the forced report task on misleading trials. Prior work has suggested that situational demands can impact the monitoring phase (Bulevich & Thomas, 2012). Warnings have also been shown to impact monitoring by reducing average confidence (Karanian et al., 2020). Thus, it could be possible that warnings could be beneficial to older adults by reducing overconfidence on misleading trials. Any improvements to monitoring could have downstream consequences on the free-report memory test, where older adults would be able to exhibit better control when given a warning. With the assumption that monitoring influences control (Nelson & Narens, 1990), it is reasonable to assume that if warnings impact monitoring, they should also impact the participants' submission criteria and confidence in the retrieval process. Therefore, I hypothesized that older adults would demonstrate improved control on misleading trials, withholding low confidence incorrect

information on the free report task, when given a warning. Support for these hypotheses could indicate that differences in monitoring and control are crucial factors underlying misinformation susceptibility in older adults.

Methods Experiment 1

Participants

I collected a total of 100 participants (100 older adults). Participants were 67.06 years old on average (SD = 4.67, range = 60-80) and mostly evenly split by gender (female, n = 52; male, n = 47; non-binary, n = 1). A power analysis was conducted in G*power using prior work on younger adults (α = .05, power = .80), which indicated a sample size of 27 participants per condition to achieve adequate power (Thomas et al., 2010). However, due to a lack of prior research with the older adult population and online administration of this paradigm, I aimed to run 50 participants per group. Data was collected using Prolific's online study recruitment platform. Participants were provided monetary compensation of $8 an hour for their time. The participants were randomly assigned to one of two warning groups: no warning and post-warning. Participants who failed attention checks or did not pass other exclusion criteria were excluded from the final analysis [failed attention check, n = 62; duplicate participants, n = 2; did not meet age criteria, n = 4, did not complete task, n = 2; spent too much time on original event (longer than 7 minutes after event video ended), n = 8; outlier (SD>2.5) in average forced report memory performance, n = 3; outlier (SD>2.5) in amount of time spent on task, n = 4]

Materials

The materials that were utilized for the video of the witnessed event, audio narration of the PEI, warning instructions and forced report memory test are taken from Karanian et al. (2020) to be able to replicate warning findings. The exact task instructions were similar but not the same. Changes were made to make the test suitable to the format of the current study's testing system (Qualtrics).

Video of Witnessed Event. An event was presented to participants during an encoding period that served as the original or witnessed event. This was a 22-minute video clip from the movie Rififi (Bezard et al., 1955). The movie is in black and white and is a film with no dialogue but contains sound effects. The clip depicts the burglary of a jewelry store, perpetrated by several assailants. Participants were instructed beforehand to pay attention to the video as they would be asked questions about the event later in the task.

Audio-narrative PEI. After watching the video, participants were provided with post-event information in the form of an audio narration of a synopsis of the witnessed event. This audio is recorded by a female speaker and consists of 115 sentences spoken at a rate of 135 to 160 words per minute that follow the chronological order of the original event. Of the 115 sentences spoken, 24 contain critical details that were tested during the memory test phase. There were 8 critical sentences of each trial type: consistent, misleading, and neutral. Each critical sentence was separated by at least 3 non-critical sentences. The information provided within these sentences either accurately described a detail from the witnessed event (*consistent*) (e.g., "Revealed at the bottom of the case is a *rope*."), inaccurately described a detail from the witnessed event (*misleading*) (e.g., "Revealed at the bottom of the case is a *towel*."), or provided a detail that was neither accurate nor inaccurate information (*neutral*) (e.g., "Revealed at the bottom of the case is a *useful object.*"). Information presented within critical sentences were counterbalanced to account for item effects with respect to trial type. All other non-critical sentences, represent an accurate summary of the original event making the majority of the audio narrative accurate.

Warning. Participants were randomly assigned to one of two warning groups. The warning is a set of instructions that inform the participant about the audio narrative (PEI). The

warning brings into question the veracity of the PEI and instructs participants to base their answers on the final test only on their memory of the witnessed event and not the audio narrative (Exact warning instructions are provided in the Appendix). The post-warning group received the warning after the audio narrative and prior to the forced report memory test. The no-warning group did not receive a warning throughout the task.

Forced Report Memory Test. Participants were given a forced report memory test after the original event and audio narrative were presented. The test consisted of 24 questions that test the participant's memory for the critical details (consistent, misleading, or neutral) from the originally witnessed event. The participants were required to respond to every question and were instructed to answer every question to the best of their abilities, even if they have to guess. Each question was a four-alternative forced-choice recognition question, presented one at a time and in chronological order (the order in which the events occurred in the video). Four alternatives were provided for each question in random order: one of them being the correct answer as was shown in the video, one of them being the misleading detail, and two highly plausible lures. For example, a question about critical information would be "What item is shown to be at the bottom of the tool case once the tools are removed?" and the options would be "Rope" (correct detail), "Towel" (misleading detail), "Knife" and "Shovel" (two plausible lures). After completing each question, participants were asked to provide a numerical rating of their confidence in their answer on a sliding scale from 0 (complete guess) to 100 (completely confident). The memory test and confidence judgments were self-paced, and participants were not able to return to a question they had completed.

Free Report Memory Test. After completing the final memory test participants were provided with instructions that they are an eyewitness that must tell only the truth and "nothing

but the truth". Following these instructions is the free report memory test, where participants were presented with each question of the forced report memory test as well as their prior answer for that question. The participants were then asked to indicate whether they believed the answers they previously provided on the forced report memory test were correct or incorrect. This served as their decision to withhold or submit a detail. This was done for each of the 24 questions about the critical details. This is the QAP methodology (Koriat & Goldsmith, 1996)

Procedure. Participants underwent an experiment where they watched a short clip from a movie and subsequently received an audio narrative summarizing the same clip. After which participants took a forced report memory test, reporting their confidence for each question, and then took a free report memory test. A diagram of the procedure can be observed in Figure 1. The experiment consisted of four primary stages: the video of the witnessed event, the audio narrative summarizing the witnessed event containing some misleading details, the forced report memory test, and the free report memory test (QAP). Participants began the experiment by first watching the video of the witnessed event on their home computer. Upon completion, the participants underwent a filler task for 10 minutes in which they were instructed to complete Sudoku puzzles. This filler task served as a distractor task to ensure that the participants would not be repeating the witnessed material and that performance on the memory tests would not be at ceiling. After the filler task they were provided with the audio narrative and directed to pay close attention to it. The duration and selected task materials were used to replicate prior work (Karanian et al. 2020). Depending on the warning group assigned, participants received a warning after the audio narrative (post-warning) or received no warning at all (no-warning). After the audio narrative and having either received no-warning or received the warning, participants took the forced report memory test, where they were required to answer each

question and provide a confidence estimate for their answer. After which, they were re-presented with those answers during a free report memory test and decide whether their previous answers are correct or not. Finally, participants completed a visual vividness questionnaire and a demographic questionnaire.

Analyses Conducted Experiment 1

The experimental design is a 2 x 2 mixed design. The variables were warning group (no-warning or post-warning) and trial type (misleading or neutral). Because my hypotheses specifically focused on misinformation susceptibility, I did not analyze consistent trials. Warning group and age are between-participants variables and trial type is a within-participants variable. The data was analyzed with a 2 x 2 mixed measures Analysis of Variance (ANOVA) in JASP (version 0.18.3) to observe the effects of warning group and trial type on various measures of memory and metamemory. Memory performance on the forced report memory test and the free report memory test was broken down by trial type (misleading and neutral). Memory performance on the forced report test was calculated by dividing the number of questions the participant answered correctly by the total number of questions within each trial type. Memory performance on the QAP, or free report test, was calculated by dividing the number of questions the participant got correct on the forced report memory test and then chose to submit on the free report memory test, by the total number of submissions on the free report memory test within each trial type. Misinformation selection was calculated by dividing the number of times a participant selected the misleading option by the total number of questions within misleading trials.

Additional analyses were conducted for metacognitive accuracy by computing a weighted confidence score for each participant for each trial type by weighing the confidence estimates at

test with the participants' accuracy on the forced report memory test to account for different numbers of observations across each trial. The method of calculating weighted calibration error was done using methods from prior work (Schraw, 2009). This was done by first binning participant confidence ratings from the forced report memory test into 12 bins with 10 equally sized bins and bins for 0 and 100. This was then compared with how frequently participants got answers with a given confidence rating correct. By comparing the value of their confidence to their accuracy, I was able to see how much error was present between participants' belief of their performance and their actual performance, thus allowing me to evaluate monitoring performance.

The submission criteria (PRC) is a value that depicts the confidence threshold above which a participant decided to submit a correct answer on the QAP (Koriat & Goldsmith, 1996). PRC was calculated using the participants' confidence scores, accuracy, and decision to submit answers on the free report. By finding the fit ratio, or ratio of hits (correct submissions above confidence) and correct rejections (inaccurate answers withheld) for every confidence estimate provided, I can find the PRC, which is the confidence value that maximizes the fit ratio. The confidence scale we utilized asked participants to respond on a 0 to 100 scale with 0 being labelled as "complete guess". However, for the sake of the PRC calculation, a lower limit of 25% is more appropriate as that is reflective of the confidence of a complete guess on a 4 alternative forced choice. Thus, prior to the calculation of the PRC's of each participant, confidence values below chance (25%) were treated as 25% for the calculation.

To observe the ability of participants to monitor and control on the free report memory test, gains and losses were calculated. Gains are the proportion of total answers that were correct

rejections and losses are the proportion of total answers that were incorrect rejections (accurate answers withheld).

A 2 (trial type: misleading, neutral) x 2 (warning: no-warning, post-warning) mixed measures ANOVA was conducted in JASP (version 0.18.3) for each dependent variable of interest, including memory performance on the free and forced report memory test, misinformation selection, confidence estimates, weighted calibration error of confidence, PRC, gains, and losses. Tukey's correction for multiple comparisons was used for all pairwise comparisons.

To quantify evidence for the absence of effects, we complemented the frequentist mixed ANOVAs with Bayesian mixed (repeated measures) ANOVAs in JASP (version 0.18.3; Wagenmakers et al., 2018a,b). Fixed factors matched the frequentist models with trial type as a within participants factor and warning group as a between participants factor, and participant entered as a random factor. Following JASP's Bayesian ANOVA implementation, model comparison evaluates the predictive performance of competing models rather than variance partitioning, and assigns equal prior model probabilities before updating these with the observed data . Unless noted otherwise, we used the default multivariate Cauchy priors for ANOVA effects as recommended for Bayesian ANOVA in JASP (Rouder et al., 2012; van den Bergh et al., 2019; Goss-Sampson, 2020; JASP Team, 2020). Because our interest is in null effects, we report Bayes factors as BF_{01} wherever the null is favored, treating Bayes factors as a continuous measure of evidence and using conventional descriptive labels only as heuristics (e.g., "anecdotal," "moderate," "strong"; Jeffreys, 1961; Lee & Wagenmakers, 2013).

Results Experiment 1

Retention: Forced report memory performance

Performance on the forced report memory test was calculated as a proportion for each subject and was calculated as the number of correct answers out of the total number of questions. This served as my measure of retention of participants. The results of the forced report memory test can be seen in Figure 2. There was a main effect of trial type on forced report memory performance (F [1,98] = 20.89, p < .001, η_p^2 = .18). Participants' retention was lower on misleading trials (M = .59) compared to neutral trials (M = .71). There was no effect of warning group on memory performance (F [1,98] = .01, p = .91 , η_p^2 = .0001 , BF_{01} = 5.44). Additionally, I did not see the expected interaction between warning group and trial type (F [1,98] = 1.69, p = .20 , η_p^2 = .02 , BF_{01} = 2.17) where warnings may have supported recognition memory for misleading trials.

Retention: Misinformation Selection

Retention can be further measured by examining misinformation selection or how often participants selected the misleading detail on the forced report test (Figure 3). As expected, I observed an effect of trial type on misinformation selection (F [1,98] = 42.18, p < .001, η_p^2 = .30) where the rate of misinformation selection was greater on misleading trials (M = .34) as compared to neutral trials (M = .17). This further supports the negative impact of misinformation on retention. There was no effect of warning group on misinformation selection (F [1,98] = .02, p = .88 , η_p^2 = .0002 , BF_{01} = 5.87) and there was no interaction between warning group and trial type for misinformation selection (F [1,98] = 1.22, p = .27 , η_p^2 = .01 , BF_{01} = 2.47).

Monitoring: Confidence

I also measured post-retrieval confidence to examine participants' monitoring during the forced report memory test (Figure 4A). Confidence was higher on misleading trials (M = .72) than on neutral trials (M = .72, F [1,98] = 18.21, p < .001, η_p^2 = .16). Warning group did not significantly impact participants' confidence estimates (F [1,98] = 2.76, p = .10 , η_p^2 = .03 , BF_{01} = 1.24) and did not interact with trial type (F [1,98] = 2.41, p = .12 , η_p^2 = .02 , BF_{01} = 1.56).

Monitoring: Calibration Error

Monitoring accuracy was examined by analyzing weighted calibration error (Figure 4B). Calibration error was greater on misleading trials (M = .23) as compared to neutral trials (M = .17; F [1,98] = 15.87, p < .001, η_p^2= .14). These findings suggest that metacognitive accuracy seems to be negatively impacted by the presence of misleading information. Warning group did not impact weighted calibration error (F [1,98] = .11, p = .11 , η_p^2 = .001 , BF_{01} = 5.34) nor interacted with trial type (F [1,98] = 1.80, p = .18 , η_p^2 = .02 , BF_{01} = 1.95).

Control: Free report memory performance

Memory performance on the free report test refers to the proportion of correct answers from the forced report test that participants chose to submit when given the option withhold answers. Forced report accuracy was calculated as a proportion based on the total number of correct answers submitted out of the total number of answers submitted. As Figure 5 illustrates, memory performance on the free report test was lower on misleading trials (M = .66) as compared to neutral trials (M = .81; F [1,98] = 29.54, p < .001, η_p^2= .23). This demonstrates that the misinformation effect persists even when the option to withhold is provided. Despite the presence of this misinformation effect, there was improvement in memory performance on the free report memory test as compared to the forced report memory test. Yet again I saw no effect of the warning group on memory performance (F [1,98] = .04, p = .84 , η_p^2 = .0004 , BF_{01} =

5.23), as well as no interaction between the warning group and trial type (F [1,98] = 1.22, p = .27 , η_p^2 = .01 , BF_{01} = 3.29). This suggests that even when given the option to withhold answers, participants were unable to use warnings to improve their memory performance.

Control: Gains and Losses

By observing the frequency of correctly withheld answers, I can see the gains in memory performance obtained through effective withholding of incorrect answers. Gains were calculated by taking the questions participants answered incorrectly on the forced report memory test and subsequently withheld on the free report memory test as a proportion of the total number of questions. On average, participants achieved gains of about 13.5%, meaning participants were able to withhold incorrect answers on the free report memory test that equated to 13.5% of the total number of questions on the memory test. Results for gains are depicted in Figure 6A. Trial type did not play a significant role in participants ability to exhibit gains (F [1,98] = 1.64, p = .20 , η_p^2 = .02 , BF_{01} = 2.92). There was no difference between warning groups in the ability to exhibit gains (F [1,98] = .10, p = .75 , η_p^2 = .001 , BF_{01} = 5.11) and no interaction between warning group and trial type (F [1,98] = .15, p = .70 , η_p^2 = .002 , BF_{01} = 6.75), suggesting that warnings did not improve older adults' ability to effectively control their retrieval.

Losses were calculated by taking number of questions participants got correct on the forced report memory test and subsequently withheld those to correct answers on the free report memory test as a proportion of the total number of questions. Participants observed an average rate of losses of about 11% (Figure 6B). Participants showed fewer losses on misleading trials (M = .08) as compared to neutral trials (M = .14; F [1,98] = 14.02, p < .001, η_p^2= .13). Warning group did not impact the rate at which people exhibited losses (F [1,98] = 1.19, p = .27 , η_p^2 =

.01 , BF_{01} = 2.95) and there was no interaction effect present (F [1,98] = 1.19, p = .28 , η_p^2 = .01 , BF_{01} = 2.81).

Control: Submission Rate

Submission rate, or number of answers submitted on the free report as a proportion of the total number of questions, allows me to better understand trends seen in losses and gains (Figure 7). Participants submitted answers more often on misleading trials (M = .79) than on neutral trials (M = .71; F [1,98] = 12.20, p < .001, η_p^2= .11). Warning group did not affect the submission rate of participants (F [1,98] = .22, p = .64 , η_p^2 = .002 , BF_{01} = 4.16) and did not interact with trial type (F [1,98] = 1.05, p = .31 , η_p^2 = .01 , BF_{01} = 2.99).

Monitoring & Control: Submission Criteria

Submission criterion refers to the level of confidence above which participants choose to submit an answer and below which they choose to withhold an answer (Figure 8). It is calculated by finding the confidence value each participant chose that maximizes the fit ratio, or ratio of correct submissions above the given confidence and inaccurate answers withheld below the given confidence. To observe the ability of participants to control on the free report memory test, gains and losses was calculated. Neither trial type (F [1,98] = 2.91, p = .09 , η_p^2 = .03 , BF_{01} = 1.65) nor warning group (F [1,98] = .53, p = .47 , η_p^2 = .005 , BF_{01} = 3.90) impact the submission criterion of participants. There was no interaction effect for submission criteria (F [1,98] = .09, p = .47 , η_p^2 = .0009 , BF_{01} = 4.64).

Discussion Experiment 1

The primary goal of Experiment 1 was to determine whether warnings and having the opportunity to withhold answers using a free report memory test could help older adults overcome the misinformation effect within a recognition memory paradigm. Contrary to my

hypotheses, evidence suggested warnings did not reduce misinformation susceptibility in older adults. Additionally, Bayesian analyses provided moderate support for the null hypothesis that performance in the warning groups did not differ for measures of retention, gains and weighted calibration error. Only anecdotal evidence was found for the null with regards to the interaction effect for these same measures other than those of control. However, older adults did effectively regulate output by withholding misleading information when given the opportunity on a free report test. Although participants improved performance on the free-report test, they continued to demonstrate the misinformation effect, with worse performance on misleading as compared to neutral trials.

Past work has shown that when older adults were provided with supportive instructions prior to testing they no longer showed deficits in memorial and metamemorial processes following exposure to misleading information (Bulevich & Thomas, 2012). Critically, Bulevich and Thomas (2012) found that this protection from misinformation in older adults was only present when the supportive instructions were paired with a cued recall test as opposed to recognition. One reason why warnings may have not been effective in Experiment 1 is because a recognition test is heavily influenced by familiarity based responding and relies less on effortful memory search (Yonelinas, 2001).

Cued recall requires individuals to search their memory to produce an answer and relies both on conscious recollective processes and familiarity. Memory search refers to the strategic, effortful process of querying one's long-term memory for specific information using retrieval cues (Bulevich & Thomas, 2012; Kahana, 2020). However, when presented with four possible choices as is done in a recognition test, individuals might opt to cut the memory search short in favor of a familiarity judgment (Yonelinas, 2001; Malmberg, 2008). When opting for a

familiarity-based judgment, older adults may be missing out on the mnemonic benefits provided by a more effortful retrieval process (Bulevich & Thomas, 2012), such as might happen in the context of warnings.

Bulevich and Thomas (2012) demonstrated that older adults showed a protection from the misinformation effect only with the supportive instructions were paired with cued recall. Thus, older adults may require warnings to be paired with a more demanding retrieval task, such as cued recall, for them to be effective. Experiment 2 aimed to test this hypothesis by investigating whether older adults could take advantage of warnings to reduce misinformation susceptibility during cued recall within the QAP methodological framework.

Experiment 2

By utilizing a cued recall task in Experiment 2, I sought to reduce reliance on familiarity-based responding and encourage more effortful retrieval, potentially enhancing older adults' ability to effectively differentiate between accurate memories and misinformation. I hypothesized that increased retrieval effort, driven by cued recall, would create optimal conditions for warnings to exert their influence. Specifically, warnings may prompt older adults to adopt a more cautious retrieval strategy by encouraging deeper consideration of source-specifying cues associated with the original event, thereby enhancing the quality of their monitoring judgments. Consequently, improved monitoring could facilitate more effective control decisions, allowing participants to set more stringent submission criteria and more accurately withhold misinformation. Thus, the cued recall format in Experiment 2 provides an ideal context to observe whether warnings can support older adults' strategic memory regulation; an effect that was not observed under conditions that favored familiarity-based responding in Experiment 1.

Methods Experiment 2

Participants

I collected a total of 100 older adult participants. Participants were 65.14 years old on average (SD = 4.80, range = 60-83) and mostly evenly split by gender (female, n = 49; male, n = 50; non-binary, n = 1). Data was collected using similar strategies to Experiment 1: Prolific's online study recruitment platform. Participants were provided monetary compensation of $8 an hour for their time. The participants were randomly assigned to one of two warning groups: no-warning and post-warning. Participants who failed the inclusion criteria were excluded from the final analysis [failed attention check, n = 90; duplicate participants, n = 2; did not meet age criteria, n = 4, did not complete task, n = 4; n = 8; outlier (SD>2.5) in average forced report memory performance, n = 3].

Materials

The same materials used in Experiment 1 were used in Experiment 2, with some minor changes to the memory tests.

Forced Report Memory Test. All details remained the same with the exception that the answer options were removed and replaced with a text entry box. For example, a question about critical information would be "What item is shown to be at the bottom of the tool case once the tools are removed?" followed by space for the participant to type in their answer. This makes the forced report memory test of Experiment 2 a cued recall task.

Free Report Memory Test. All details remained the same with the exception that instead of answer options being re-presented, participants were re-presented with what they had typed in for their answer on the forced report memory test.

Analyses Conducted Experiment 2

I utilized the same analyses planned for Experiment 1 for the resulting data from Experiment 2. The only difference is that misleading selection has been switched with misleading production. As the forced report memory test was a cued recall task in Experiment 2 with no options to select from, misinformation here was produced as opposed to selected. With that in mind, as the misleading details were plausible answers, even without ever having received misleading details, participants can still inadvertently produce the misleading detail. This is why it is possible to observe a non-zero value of misinformation production even on neutral trials.

The same 2 (trial type: misleading, neutral) x 2 (warning group: no-warning, post-warning) mixed measures ANOVAs were conducted on the variables of memory performance on the free and forced report memory test, misinformation production, confidence estimates, weighted calibration error of confidence, PRC, gains, and losses. Tukey's correction for multiple comparisons was used for all pairwise comparisons once again.

The same 2 (trial type: misleading, neutral) x 2 (warning group: no-warning, post-warning) mixed measures Bayesian ANOVAs were utilized for any null findings for the variables of memory performance on the free and forced report memory test, misinformation production, confidence estimates, weighted calibration error of confidence, PRC, gains, and losses.

Results Experiment 2

Results for Experiment 2 were calculated and analyzed in the same manner as was done for Experiment 1, the only difference being that the forced report memory test was a cued recall test. As a result of this change, misinformation selection has been changed into misinformation production as a measure for misinformation susceptibility where production refers to the participant writing the misleading detail as their answer.

Retention: Forced report memory performance

As in Experiment 1, retention, in the form of the forced report memory test performance, was calculated as the proportion of the total number of questions on the forced report that the participant answered correctly (Figure 9). I found a main effect of trial type on memory performance, (F [1,98] = 8.36, p = .0047, η_p^2 = .08) where participants showed reduced performance on misleading trials (M = .45) compared to neutral trials (M = .52). I did not find an effect of the warning group (F [1,98] = .07, p = .80 , η_p^2 = .0007 , BF_{01} = 4.77), or the predicted interaction between warning group and trial type (F [1,98] = .54, p = .46 , η_p^2 = .006 , BF_{01} = 3.69).

Retention: Misinformation Production

Misinformation production was calculated the number of answers produced on the forced report memory test that matched the misleading detail as a proportion of the total number of questions. As was seen in Experiment 1, I observed an effect of trial type on misinformation production (F [1,98] = 89.46, p < .001, η_p^2 = .48) where participants produced the misinformation more often on average on misleading trials (M = .32) as compared to neutral trials (M = .10) (see Figure 10). As was observed in Experiment 1, there was neither an effect of warning group (F [1,98] = 2.03, p = .16 , η_p^2 = .02 , BF_{01} = 3.76) nor an interaction effect of

warning group with trial type (F [1,98] = .69, p = .41 , η_p^2 = .007 , BF_{01} = 3.41) on misinformation production.

Monitoring: Confidence and Calibration Error

Confidence was calculated as the average confidence participants reported at the forced report memory test across each question. Unlike in Experiment 1, I did not find an effect of trial type on memory confidence (F [1,98] = 2.03, p = .16 , η_p^2 = .02 , BF_{01} = 2.59) (see Figure 11A). The effect of warning was also non-significant (F [1,98] = .16, p = .69 , η_p^2 = .002 , BF_{01} = 2.67). I also found no significant interaction between warning group and trial type (F [1,98] = .68, p = .41 , η_p^2 = .007 , BF_{01} = 3.33). Similar to Experiment 1, participants had a greater calibration error on misleading trials (M = .25) as compared to neutral trials (M = .21; F [1,98] = 8.34, p = .0048, η_p^2= .04) (See Figure 11B). Again, warning group did not impact calibration error (F [1,98] = 1.37, p = .24 , η_p^2 = .01 , BF_{01} = 3.23) and there was no interaction effect present (F [1,98] = .002, p = .97 , η_p^2 = .00002 , BF_{01} = 4.84).

Control: Free report memory performance

Free report memory performance was calculated as the number of questions that were answered correctly on the forced report memory test and were subsequently submitted as a proportion of the total number of answers submitted. Consistent with Experiment 1, I observed significantly lower memory performance on misleading trials (M = .52) as compared to neutral trials (M = .65; F [1,98] =14.62, p < .001, η_p^2= .13), when participants were allowed to withhold responses. I saw no effect of the warning group on memory performance (F [1,98] = .03, p = .86 , η_p^2 = .0003 , BF_{01} = 5.30), as well as no interaction between the warning group and trial type (F [1,98] = 2.21, p = .14 , η_p^2 = .02 , BF_{01} = 1.73). These results suggest that even when the

increased retrieval effort of cued recall is combined with the option to withhold, this is insufficient in decreasing older adults' misinformation susceptibility.

Control: Gains and Losses

Gains were calculated as the number of questions which participants answered incorrectly on the forced report memory test and were subsequently withheld as a proportion of the total number of questions. Losses were similarly calculated as the number of questions which participants answered correctly on the forced report memory test and were subsequently withheld as a proportion of the total number of questions. Compared to Experiment 1, I saw participants achieving a higher level of gains (24%) and reduced losses (9%). I saw no effects of either trial type (F [1,98] = .51, p = .47 , η_p^2 = .005 , BF_{01} = 4.93), warning group (F [1,98] = .42, p = .52 , η_p^2 = .004 , BF_{01} = 4.18) on gains as well as no interaction effect (F [1,98] = .03, p = .87 , η_p^2 = .0003 , BF_{01} = 4.53). I also found no effects of either trial type (F [1,98] = .20, p = .66 , η_p^2 = .002 , BF_{01} = 5.96), warning group (F [1,98] = 1.25, p = .26 , η_p^2 = .01 , BF_{01} = 2.87), or interaction effect (F [1,98] = .64, p = .43 , η_p^2 = .007 , BF_{01} = 3.10) on losses.

Submission Rate and Submission Criterion

Submission rate is calculated as the proportion of the total number of questions that were submitted on the free report memory test. Submission criterion is calculated as the confidence value at which the fit ratio, or ratio of correct answered submitted above the confidence estimate to incorrect answers withheld below the confidence estimate, is maximized. Analyses of submission rate showed no effects of trial type (F [1,98] = .93, p = .34 , η_p^2 = .01 , BF_{01} = 4.21) or warning group (F [1,98] = .004, p = .95 , η_p^2 = .00004 , BF_{01} = 4.17) nor an interaction between trial type and warning group (F [1,98] = .10, p = .10 , η_p^2 = .001 , BF_{01} = 4.13). I did not see any effects of trial, similar to that of Experiment 1, on submission criteria (F [1,98] = 2.19, p

= ..14 , η_p^2 = .02 , BF_{01} = 2.42) nor an interaction effect (F [1,98] = 2.29, p = .13, η_p^2= .02, BF_{01} = 1.62). However, I did observe a main effect of warning group (F [1,98] = 5.92, p = .017, η_p^2= .06) where participants in the post-warning group (M = .76) had a higher submission criterion than those in the no-warning group (M = .67). These results suggest that warnings make individuals more selective when deciding what answers to submit.

Discussion Experiment 2

Despite increasing retrieval effort by shifting to a cued-recall format, warnings did not meaningfully influence retention, monitoring, or control in our older adult sample. Warnings also did not improve performance on the free report test. Bayesian analyses further provided moderate evidence in favor of the null hypothesis for warning in most measures of retention monitoring and control, except for confidence, like in Experiment 1. Furthermore, moderate support for null effects of interaction were present for retention, monitoring and control except for confidence, free report memory performance and losses. While the overconfidence on misleading trials that was found in Experiment 1 was reduced in Experiment 2, calibration error was still greater on misleading trials as compared to neutral trials. Importantly, in Experiment 2 warnings raised participants' submission criterion, indicating that one aspect of metamemorial control was positively impacted. However, another measure of control, participants' ability to achieve gains through withholding incorrect information, remained unaffected by warnings. Finally, as in Experiment 1, misleading details continued to impair retention and misinformation production rates were high. These results suggest that although warnings alone were insufficient in significantly reducing misinformation susceptibility, the observed increase in selectivity (submission criterion) indicates that older adults can still adaptively respond to situational

demands, reinforcing the importance of examining retrieval conditions that might best support memory accuracy and metamemorial processes in older adults.

General Discussion

A large body of work has found that the memorial and metamemorial processes of retention, monitoring and control are impaired by the presence of misinformation in older adults (Kelley & Sahakyan, 2001; Mitchel et al., 2003; Roediger & Geraci, 2007; Bulevich & Thomas, 2012). A warning against misleading information has been showed to be beneficial in reducing the misinformation effect in younger adults (Blank & Launay, 2014; Thomas et al., 2010). Despite evidence that warnings can reduce misinformation susceptibility in younger adults, it remains unclear whether such warnings similarly support older adults' retention, monitoring, and control processes. The current study aimed to address this question by examining the efficacy of warnings in supporting these memory and metamemorial processes within an eyewitness misinformation paradigm specifically in older adults.

In line with prior work (Kelley & Sahakyan, 2001; Bulevich & Thomas, 2012), misinformation impaired both retention and monitoring accuracy in older adults across both experiments, as demonstrated by decreased memory performance on the forced report test and increased calibration error on misleading trials. However, unlike in younger adults, warnings proved to be ineffective at improving memory or metamemorial processes. While Bayesian analyses did not provide strong evidence in favor of the null hypothesis of no difference between warning groups or the presence of an interaction, warnings did not have a robust effect on older adults' retention, monitoring, or control across either experiment. Importantly, older adults demonstrated an intact ability to strategically regulate memory output in both Experiments as demonstrated by similar levels of gains in accurate memory output on neutral and misleading

trials. This novel finding differs from Bulevich and Thomas (2012) who found this pattern of results only when participants were given supportive instructions paired with a cued recall test. This suggests that older adults can effectively regulate their memory output even without explicit supportive instructions, provided that task demands do not exceed their available cognitive resources; thus, highlighting the importance of considering cognitive load and retrieval conditions when developing interventions to support eyewitness accuracy in older populations.

Why Warnings May Not Have Worked: Encoding Deficit

There are several plausible explanations for why warnings may not have impacted memory or metamemory processes in the current study. One possibility is that older adults may not have encoded the warnings or paid attention to it. This possibility is supported by the absence of significant differences in forced report memory performance, misinformation selection, and calibration error between the warning groups across both experiments, indicating that warnings did not significantly alter initial retrieval processes. However, older adults did show an increase in their submission criterion in the presence of a warning, suggesting some impact of warning. Importantly, the warning's impact on monitoring did not have downstream consequences for control, or overall memory accuracy. This suggests that the reasons warnings were not effective is likely not due to older adults having not encoded the warning.

Why Warnings May Not Have Worked: Source Monitoring Deficit

Warnings have been shown to bias retrieval processes toward the accurate source of information in younger adults, supporting accurate source memory through mechanisms such as sensory cortical reinstatement (Karanian et al., 2020). Specifically, younger adults who received warnings exhibited increased neural reactivation of sensory regions corresponding to the accurate source of information, facilitating correct memory retrieval. Source misattribution error

is thought to be one of the underlying factors associated with the misinformation effect (Lindsay & Johnson, 1989). Older adults, however, demonstrate pronounced difficulties with source monitoring tasks due to age-related declines in strategic retrieval and attentional control processes necessary for accurately distinguishing between multiple sources of information (Mitchell & Mather, 2003). Consequently, older adults' diminished capacity for effective source monitoring could prevent them from utilizing warnings as younger adults do. If warnings primarily exert their protective influence by promoting accurate cortical reinstatement and enhanced scrutiny of memory sources during retrieval, older adults' deficits in these source monitoring processes may severely limit the efficacy of warnings, resulting in continued susceptibility to misinformation.

Though the inability to use warnings could stem from deficits in the ability to source monitor, research utilizing an explicit task to encourage source memory suggests otherwise. Multhaup et al. (1999) tested older adults in an eyewitness memory paradigm and provided an explicit source monitoring task at retrieval to some of the participant. Older adults who received the source monitoring task were less susceptible to misinformation. This suggests that older adults are able to accurately monitor the source of retrieved memories. This also showcased older adults' ability to use strategies during retrieval when explicitly told to do so. Multhaup et al.'s (1999) source monitoring task was a behavioral task that demanded explicit source monitoring. On the other hand, the warning utilized in the current study warns participants about the exposure to misinformation and advises them to retrieve from the original source. In comparison, an explicit task encouraged older adults to partake in a more effortful retrieval. Thus, the reason why warnings might not have worked could be insufficient retrieval effort.

Why Warnings May Not Have Worked: Insufficient Retrieval Effort

One critical factor influencing older adults' vulnerability to misinformation is retrieval effort. Increased retrieval effort has previously been shown to improve older adults' retention, monitoring accuracy, and control by fostering deeper, more exhaustive memory searches, thereby allowing better differentiation between accurate and misleading details (Bulevich & Thomas, 2012; Robinson & Johnson, 1996). Older adults' memory retrieval is often influenced by familiarity-based responding, which biases retention, monitoring, and subsequent control decisions. Familiarity can artificially inflate confidence in incorrect details, impairing older adults' ability to distinguish between accurate memories and misinformation (Chua et al., 2009; Yonelinas, 2001). Indeed, in Experiment 1, older adults showed overconfidence in misleading trials, likely due to this familiarity bias inherent in recognition-based memory tests.

In Experiment 2, the shift from recognition to cued recall successfully reduced reliance on familiarity, evident through a reduction in overconfidence on misleading trials. Despite this, warnings did not significantly improve older adults' retention or control performance, although they increased the submission criterion, indicating greater caution. Thus, while older adults became more selective when warned, their underlying retention and monitoring processes did not improve, suggesting that warnings alone did not sufficiently enhance retrieval effort.

Supportive instructions used by Bulevich and Thomas (2012) explicitly aimed to increase retrieval effort through detailed guidance on how to search memory and manage retrieved details, substantially improving older adults' resistance to misinformation. In contrast, the warnings used in the present study lacked explicit retrieval strategies, merely informing participants of possible misinformation without guiding them on memory search techniques. This general and indirect approach likely failed to trigger the deep, source-specific retrieval processes crucial for older adults. Indeed, Multhaup et al. (1999) demonstrated that older adults effectively

resisted misinformation when explicitly prompted to engage in source monitoring, highlighting that older adults benefit significantly from explicit, strategy-focused instructions. Thus, the general warnings in the present study may have been inadequate because they did not explicitly prompt older adults to engage the retrieval effort necessary to effectively discriminate between accurate memories and misinformation.

Critically, Bulevich and Thomas (2012) found that older adults only demonstrated reduced misinformation susceptibility when both supportive instructions and cued recall tasks were provided together. Neither supportive instructions nor cued recall alone was sufficient in producing this protective effect, underscoring the necessity of combining explicit retrieval guidance with conditions promoting high retrieval effort to enhance older adults' ability to resist misinformation effectively. Thus, the absence of explicit supportive instructions in the current study likely limited the effectiveness of warnings, even under the conditions of increased retrieval effort provided by cued recall in Experiment 2. Consistent with this interpretation, the Bayesian model comparisons favored the simpler models without warning for most outcomes. This suggests that a general warning, one without explicit, strategy-level support, was unlikely to have altered the underlying retrieval processes in older adults. Future research should explore whether integrating explicit source-monitoring instructions into warnings can better support older adults' memory and metamemory performance in the face of misinformation.

Gains in Strategic Regulation

While older adults were unable to use warnings effectively, critically, the current study found that they still demonstrated substantial gains on misleading trials on the free report test when given the opportunity to withhold answers. This contrasts with prior findings where older adults typically showed fewer gains on misleading compared to neutral trials (Bulevich &

Thomas, 2012; Kelley & Sahakyan, 2001) and indicates that older adults' strategic regulation abilities can remain intact in certain circumstances, suggesting that the presence of misinformation itself does not inherently diminish older adults' ability to control their memory output. The divergence in findings between the current study and prior research (Bulevich & Thomas, 2012) may be attributed to methodological differences. Specifically, the present study utilized a shorter, 22-minute video, whereas Bulevich and Thomas employed a significantly longer, 40-minute video. Additionally, the current study featured an auditory post-event narrative as opposed to a written one and contained fewer critical details and memory test questions. These variations presumably reduced overall task demands.

Indeed, Castel's value-directed remembering research demonstrates that healthy older adults can efficiently allocate limited cognitive resources to prioritize important information when task demands are manageable (Castel et al., 2002; Castel, 2008). Recent work by Murphy and Castel's (2024) confirms that older adults' strategic regulation benefits from conditions with lower cognitive demands. Conversely, tasks that heavily tax working memory or require complex cognitive operations and extensive self-initiated processing tend to diminish older adults' ability to effectively engage in strategic regulation (Craik & Bosman, 1992; Salthouse, 2010; Murphy, 2024). Thus, with fewer cognitive demands, older adults in the current study may have been better able to strategically allocate their limited cognitive resources, enabling improved regulation and resulting in substantial gains even in the presence of misinformation.

Limitations

The current study has several limitations related primarily to generalizability and methodological constraints inherent in online data collection. First, the online administration of

the experiment limits the control over participants' testing environments. Variability in testing conditions, such as distractions, inconsistent audio quality during the PEI narration, and differences in participants' technological proficiency, could contribute to noise in the data that would be less prominent in controlled laboratory settings. Additionally, despite rigorous exclusion criteria and attention checks, the high rate of participant exclusion in both experiments suggests that online participants might engage with tasks differently than participants in traditional laboratory-based research, potentially impacting the representativeness of these findings.

Furthermore, the sample itself, drawn exclusively from the Prolific online participant pool, may limit the demographic and cognitive diversity of the study population. Older adults who regularly participate in online studies might differ systematically from the broader population in terms of cognitive health, familiarity with technology, and engagement levels, potentially inflating or deflating the study's observed effects. Thus, caution must be exercised when attempting to generalize the findings to broader populations of older adults or eyewitnesses.

Another limitation pertains to the chosen stimuli. The video used for the witnessed event was a black-and-white, dialogue-free film segment adapted from Karanian et al. (2020). Such a video may lack ecological validity relative to typical eyewitness scenarios involving rich visual and auditory contextual details. The absence of dialogue and color could reduce encoding depth and event salience, potentially impacting the retrieval processes central to resisting misinformation. While this decision was methodologically justifiable for replicability, future research should explore more ecologically valid stimuli to enhance applicability to real-world eyewitness contexts.

Conclusion

The present study investigated how warnings influence older adults' memory, monitoring, and control processes in the presence of misinformation. Across two experiments, results consistently demonstrated the persistence of the misinformation effect among older adults, negatively affecting memory accuracy, monitoring calibration, and control of memory output. Contrary to expectations, warnings alone did not significantly mitigate the negative effects of misleading post-event information. Despite this, older adults exhibited intact strategic regulation by withholding incorrect information when provided with the opportunity. The findings align with prior work by Bulevich and Thomas (2012) emphasizing the importance of retrieval effort in supporting older adult memory and metamemory. The present study showcases that warnings without explicit retrieval strategies are insufficient for older adults. Overall, the results underscore the complexity of misinformation susceptibility in older eyewitnesses, highlighting the necessity of explicit retrieval support rather than general warnings alone to improve memory accuracy and metamemory judgments in real-world scenarios.

Figures

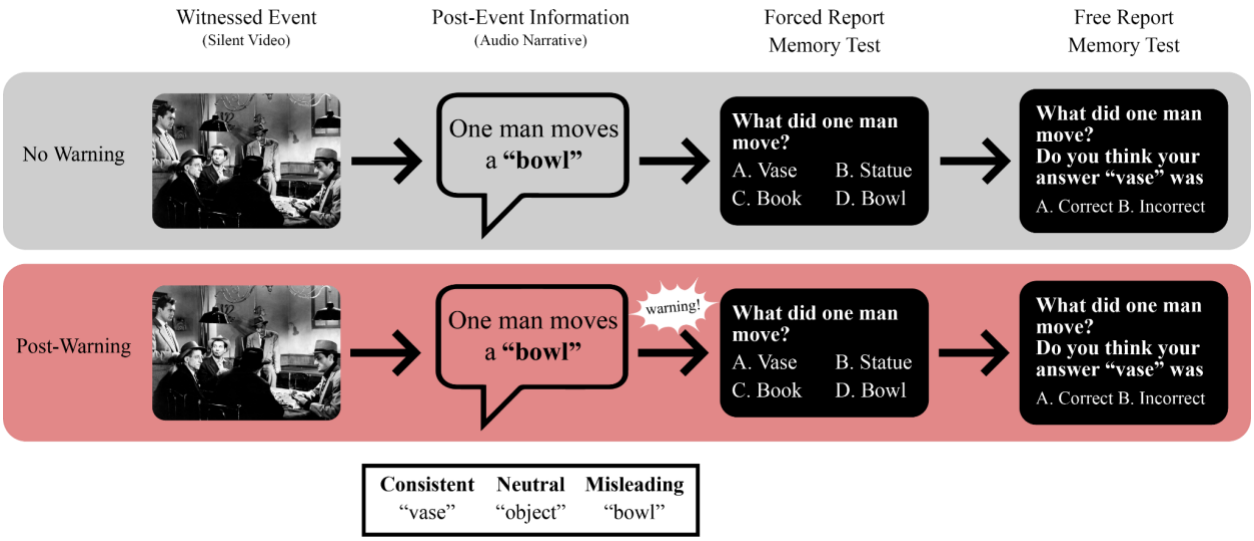

Figure 1. The eyewitness paradigm that was utilized in Experiment 1. Participants watched a silent video depicting the witnessed event and then listen to an audio narrative containing post-event information with critical details that are either consistent, misleading, or neutral with respect to the original event. After the audio narrative, participants completed a forced report recognition memory test for their memory of the witnessed event. Participants were then given the choice to withhold any answers they did not believe they answered correctly on the free report memory test. Importantly, half of the participants were warned about the veracity of the post-event information after the audio narrative (post-warning). The remaining participants received no-warning (no-warning).

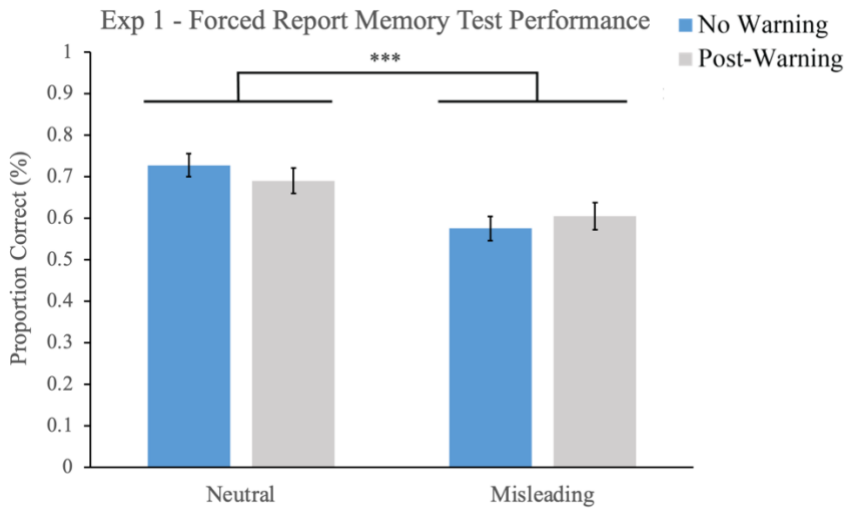

Figure 2. Proportion correct on the forced report memory test by trial type and warning group. Proportion correct refers to the total number of questions responded to correctly on the forced report memory test split by trial type. Error bars indicate between participant standard error.

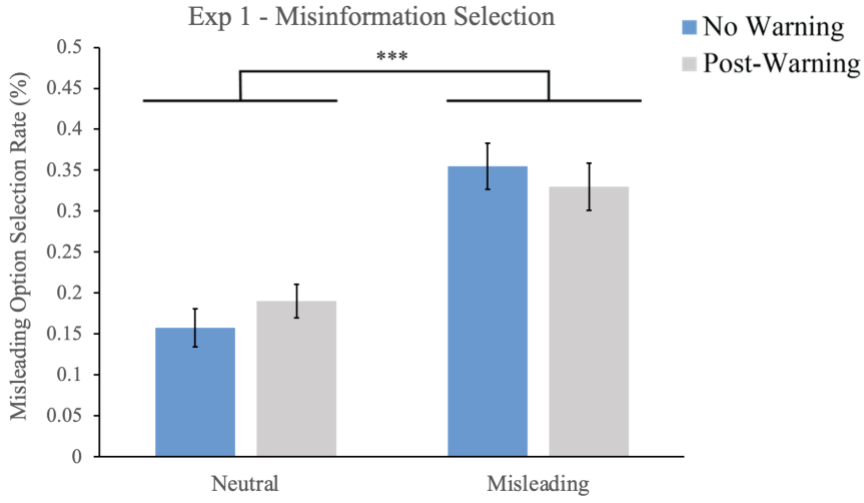

Figure 3. Misinformation selection by trial type and warning group. Misleading option selection rate refers to the proportion of total questions on the forced report test on which the misleading option was selected among the 4 forced choices. Error bars indicate between participant standard error.

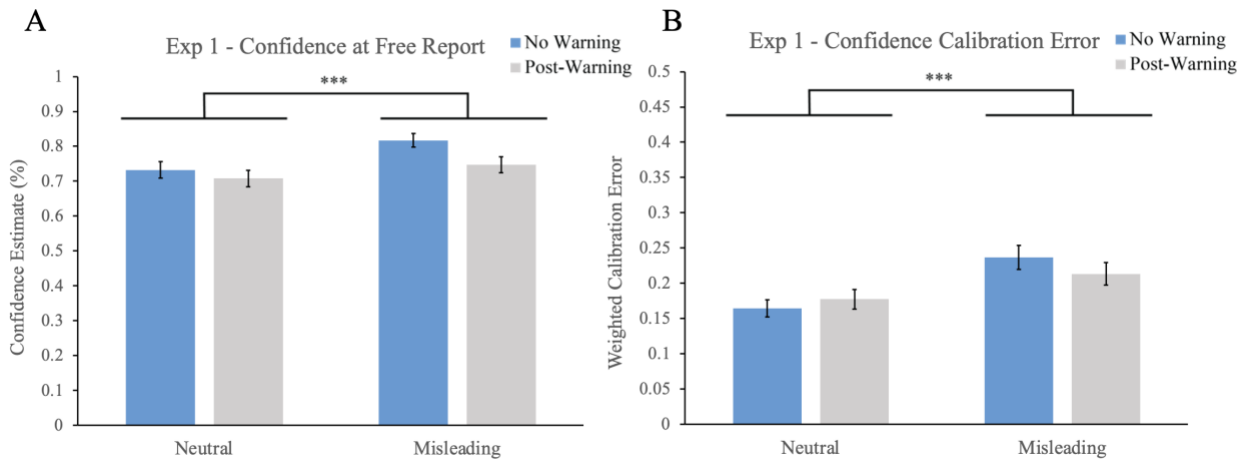

Figure 4. (A) Confidence judgments made during the forced report test after every question split by trial type and warning group. Confidence estimate refers to the likelihood the participant believed their response was correct (B) Error in calibration of participants' confidence estimate by trial type and warning group. Weighted calibration error refers to the error on calibration between participants' confidence and memory performance weighted by the frequency of responses. Error bars indicate between participant standard error.

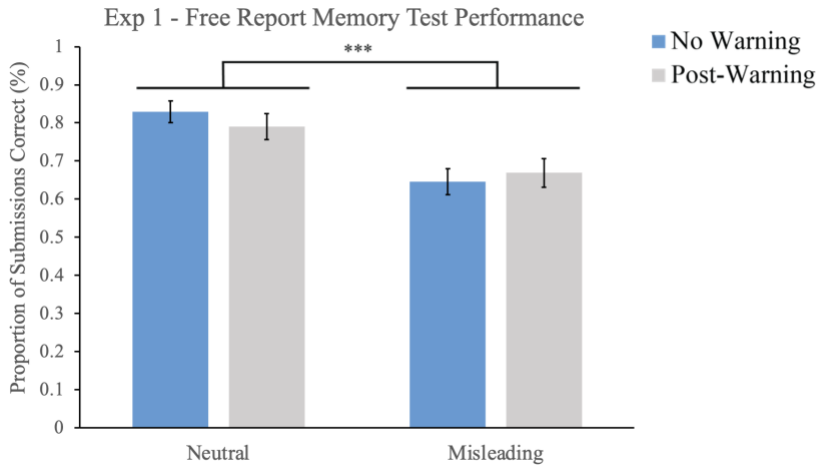

Figure 5. Memory performance on the free report memory test by trial type and warning group. Memory performance refers to the total number of questions responded to correctly on the forced

report memory test and was chosen to be submitted on the free report memory test, split by trial type. Error bars indicate between participant standard error.

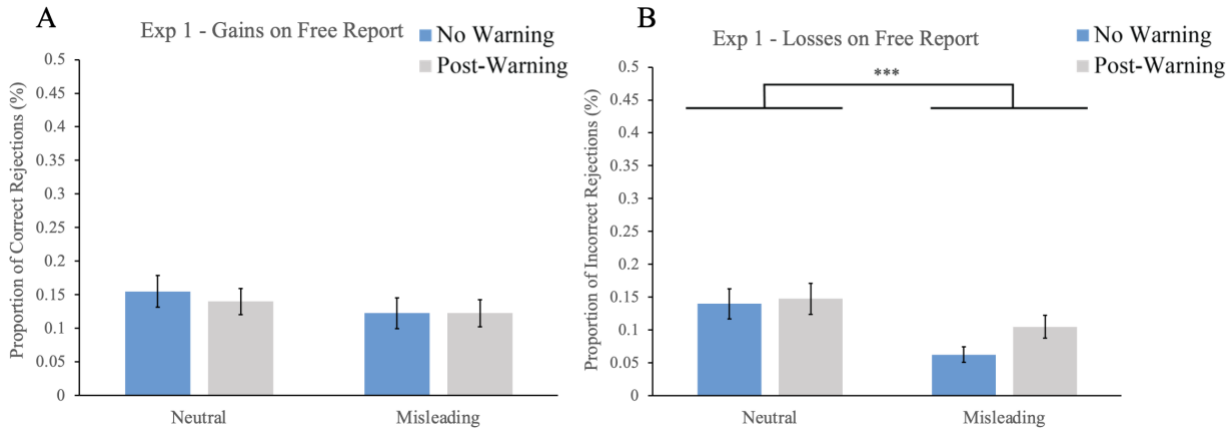

Fig 6. Control exhibited by participants by withholding seen in gains and losses by trial type and warning group. (A) Results for gains. Proportion of correct rejections refers to the total number of questions participants got incorrect on the forced report memory test and subsequently withheld the answer on the free report test. (B) Results for losses. Proportion of incorrect rejections refers to the total number of questions participants got correct on the forced report memory test and subsequently withheld the answer on the free report test.

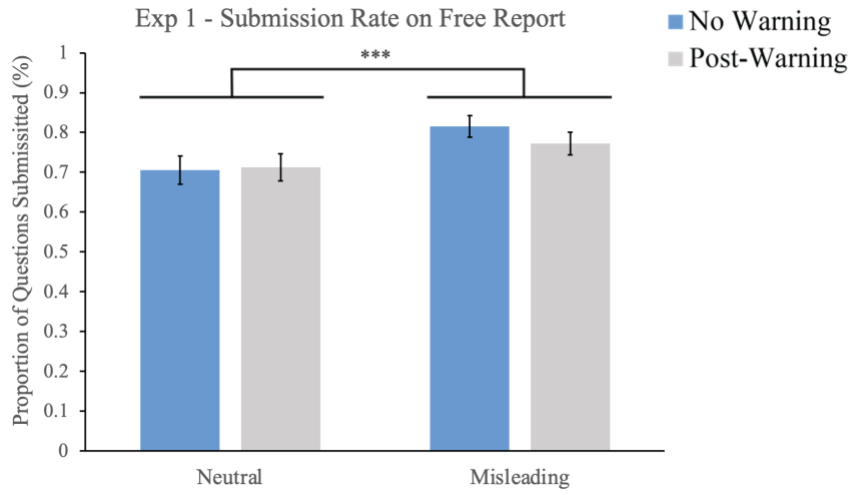

Fig 7. Submission rate of participants split by trial type and warning group. Proportion of questions submitted refers to proportion of questions of each trial type participants chose to submit on the free report memory test. Error bars indicate between participant standard error.

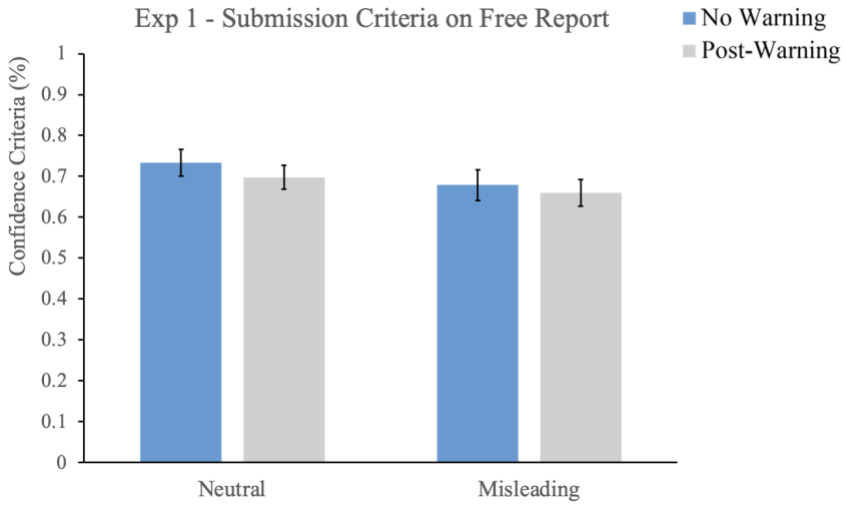

Fig 8. Submission criteria of participants split by trial type and warning group. Confidence criteria refers to the confidence level above which participants chose to submit an answer and below which chose to withhold an answer on the free report memory test. Error bars indicate between participant standard error.

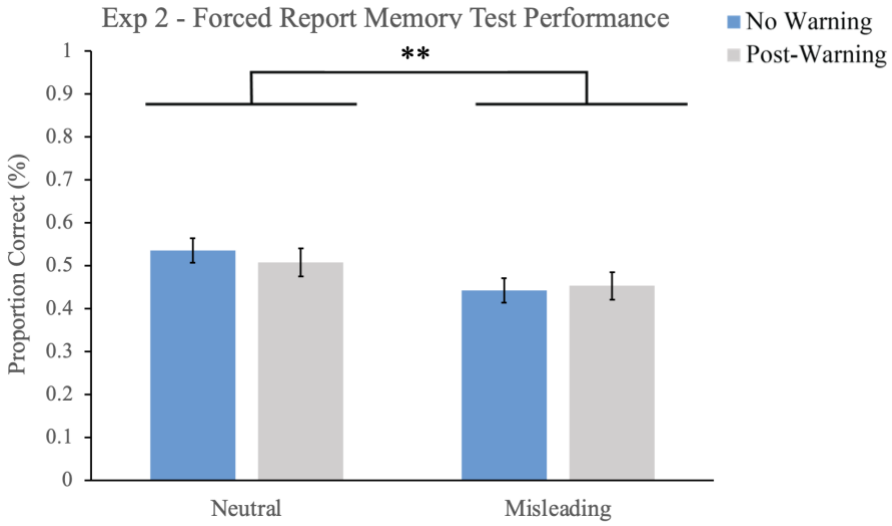

Figure 9. Proportion correct on the forced report memory test by trial type and warning group. Proportion correct refers to the total number of questions responded to correctly on the forced report memory test split by trial type. Error bars indicate between participant standard error.

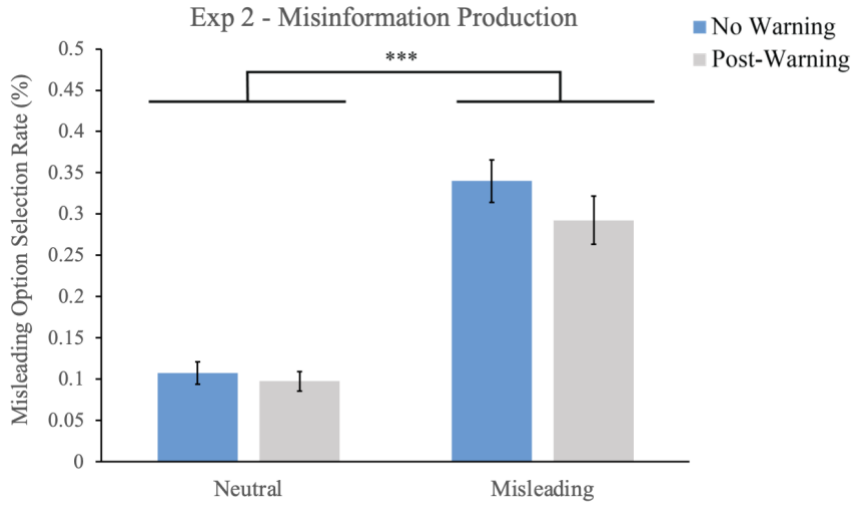

Figure 10. Misinformation production by trial type and warning group. Misleading detail rate refers to the proportion of total questions on the forced report test on which the misleading detail was produced as an answer on the cued. Error bars indicate between participant standard error.

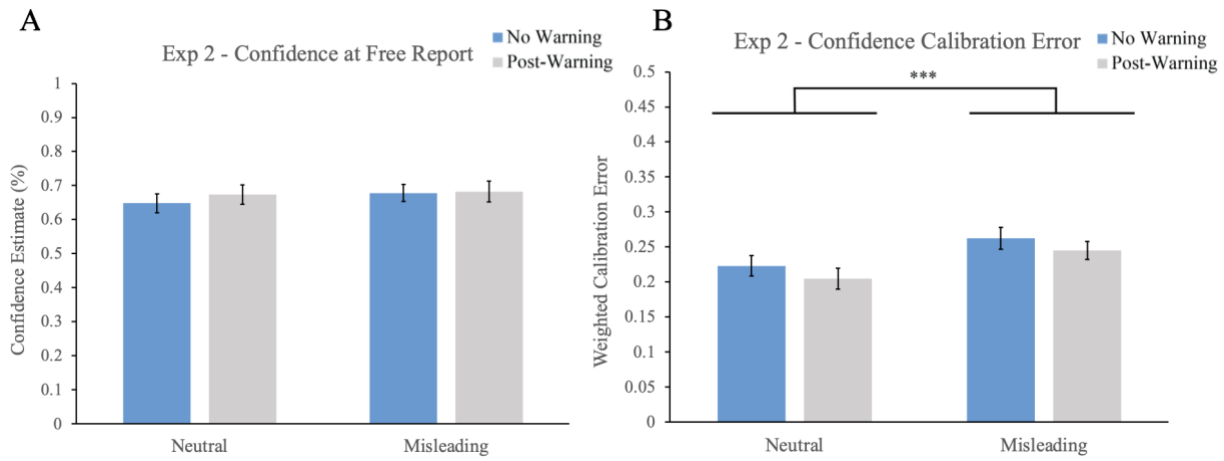

Figure 11. (A) Confidence judgments made during the forced report test after every question split by trial type and warning group. Confidence estimate refers to the likelihood the participant believed their response was correct (B) Error in calibration of participants' confidence estimate by trial type and warning group. Weighted calibration error refers to the error on calibration between participants' confidence and memory performance weighted by the frequency of responses. Error bars indicate between participant standard error.

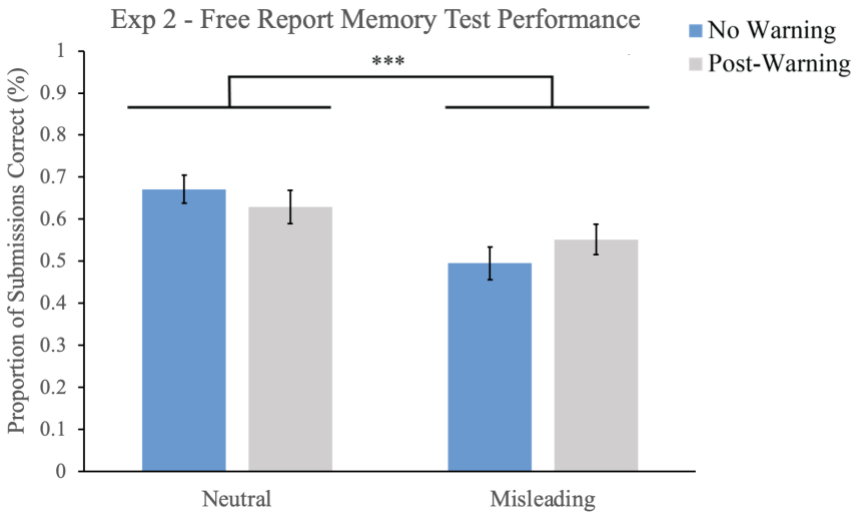

Figure 12. Memory performance on the free report memory test by trial type and warning group. Memory performance refers to the total number of questions responded to correctly on the forced

report memory test and was chosen to be submitted on the free report memory test, split by trial type. Error bars indicate between participant standard error.

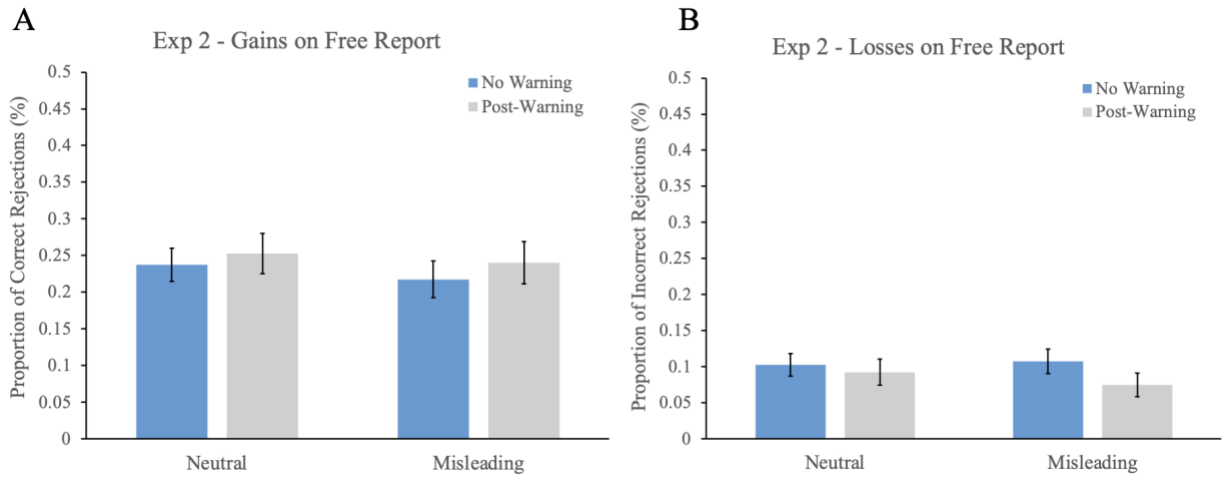

Fig 13. Control exhibited by participants by withholding seen in gains and losses by trial type and warning group. (A) Results for gains. Proportion of correct rejections refers to the total number of questions participants got incorrect on the forced report memory test and subsequently withheld the answer on the free report test. (B) Results for losses. Proportion of incorrect rejections refers to the total number of questions participants got correct on the forced report memory test and subsequently withheld the answer on the free report test. Error bars indicate between participant standard error.

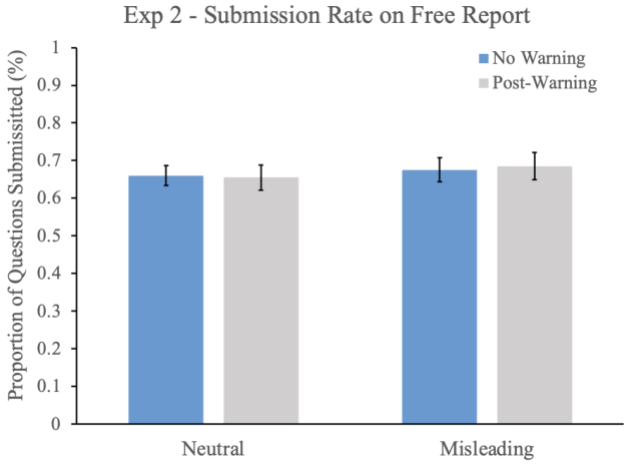

Fig 14. Submission rate by participants by trial type and warning group. Proportion of questions submitted refers to the number of questions on the free report memory test for which the participant chose to submit an answer by declaring it to be “Correct”. Error bars indicate between participant standard error.

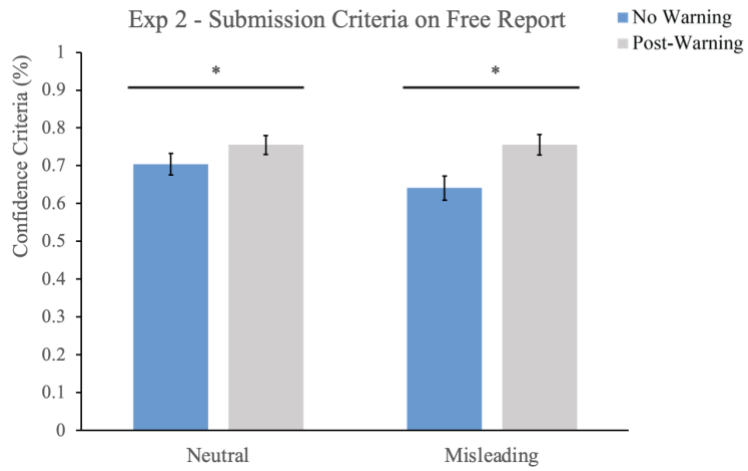

Fig 15. Submission criteria of participants split by trial type and warning group. Confidence criteria refers to the confidence level above which participants chose to submit an answer and below which chose to withhold an answer on the free report memory test. Error bars indicate between participant standard error.

Appendix

Warning instructions provided to participants in the no warning and post-warning group in both Experiment 1 and 2:

No Warning Group

Prior to playing the audio narrative:

> *"You will now be presented with an audio narrative or synopsis of the video you watched. The next button will appear after the narrative is over.*
> *Please listen to the narrative carefully."*

After playing the audio narrative:

> *"Now you will be given a series of questions to answer based on the video you just watched. In the video you had just watched you saw several crimes being committed. Therefore, think of yourself as an eyewitness to the crime.*
> *As an eyewitness, it is important that you answer the following questions as accurately as possible. Additionally, we want as many details as you can possibly remember. Therefore, we ask that you answer EVERY question.*
> *If you are unsure about the answer to a question, please make your best guess. It's important to guess even if you have no idea.*
> *After each answer, please provide your confidence in that answer on a scale of 0 (not confident at all) to 100 (extremely confident). If you were very uncertain about an answer, make sure to give that a lower confidence rating. We encourage you to use the entire scale for your answers."*

Post-Warning Group

Prior to playing the audio narrative:

> *"You will now be presented with an audio narrative or synopsis of the video you watched. The next button will appear after the narrative is over.*
> *Please listen to the narrative carefully."*

After playing the audio narrative:

> *"Now you will be given a series of questions to answer based on the video you just watched. In the video you had just watched you saw several crimes being committed. Therefore, think of yourself as an eyewitness to the crime.*
> ***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.***
> *As an eyewitness, it is important that you answer the following questions as accurately as possible. Additionally, we want as many details as you can possibly remember. Therefore, we ask that you answer EVERY question.*
> *If you are unsure about the answer to a question, please make your best guess. It's important to guess even if you have no idea.*
> *After each answer, please provide your confidence in that answer on a scale of 0 (not confident at all) to 100 (extremely confident). If you were very uncertain about an*

answer, make sure to give that a lower confidence rating. We encourage you to use the entire scale for your answers.

Bibliography

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: I. Learning, memory and thinking.* W. H. Freeman.

Barzykowski, K., Staugaard, S., & Mazzoni, G. (2021). Retrieval effort or intention: Which is more important for participants' classification of involuntary and voluntary memories?. *British journal of psychology.* https://doi.org/10.1111/bjop.12498.

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition, 3*(2), 77–88. https://doi.org/10.1037/h0101798

Bornstein, B. H. (1995). Memory processes in elderly eyewitnesses: What we know and what we don't know. *Behavioral Sciences & the Law,* 13, 349–363. doi:10.1002/bsl.2370130303

Buckner, R., Koutstaal, W., Schacter, D., Wagner, A., & Rosen, B. (1998). Functional–Anatomic Study of Episodic Retrieval Using fMRI I. Retrieval Effort versus Retrieval Success. *NeuroImage*, 7, 151-162. https://doi.org/10.1006/nimg.1998.0327.

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

Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *Skill and strategy in memory use* (pp. 225–270). Elsevier Academic Press.

Castel, A. D., Humphreys, K. L., Lee, S. S., Galván, A., Balota, D. A., & McCabe, D. P. (2011). The development of memory efficiency and value-directed remembering across the life span: a cross-sectional study of memory and selectivity. *Developmental psychology*, *47*(6), 1553–1564. https://doi.org/10.1037/a0025623

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural basis for recognition confidence in younger and older adults. *Psychology and aging*, 24(1), 139–153. https://doi.org/10.1037/a0014029

Craik, F. I. M., & Bosman, B. A. (1992). Age-related changes in memory and learning. In H. Bouma & J. A. M. Graafmans (Eds.), *Gerontechnology* (pp. 79–92). IOS Press.

Daniels, K. A., Toth, J. P., & Hertzog, C. (2009). Aging and recollection in the accuracy of judgments of learning. *Psychology and aging*, 24(2), 494–500. https://doi.org/10.1037/a0015269

DeCaro, R., Thomas, A.K. Prompting retrieval during monitoring and self-regulated learning in older and younger adults. *Metacognition Learning* 15, 367–390 (2020). https://doi.org/10.1007/s11409-020-09230-y

Dodson, C. S., & Krueger, L. E. (2006). I misremember it well: Why older adults are unreliable eyewitnesses. *Psychonomic Bulletin & Review,* 13(5), 770–775.

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, 22(1), 122–133.

Dunlosky, J., & Bjork, R.A. (2013). Handbook of Metamemory and Memory. Psychology Press.

Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 11–28). Psychology Press.

Earles, J.L., Kersten, A., Berlin Mas, B., & Miccio, D.M. (2004). Aging and memory for self-performed tasks: effects of task difficulty and time pressure. *The journals of gerontology. Series B, Psychological sciences, and social sciences, 59 6*, P285-93 .

Ford, J.H., Rubin, D.C., & Giovanello, K.S. (2014). Effects of task instruction on autobiographical memory specificity in young and older adults. *Memory*, 22, 722 - 736.

Goss-Sampson, M. A. (2020). *Bayesian inference in JASP 0.12.2: A guide for students* (Version 0.12.2) [PDF manual]. OSF. https://doi.org/10.17605/OSF.IO/CKNXM

Greene, E., & Loftus, E. F. (1984). Solving the eyewitness problem. *Behavioral Sciences & the Law, 2*(4), 395–406. https://doi.org/10.1002/bsl.2370020406

Halamish, V., McGillivray, S., & Castel, A. D. (2011). Monitoring one's own forgetting in younger and older adults. *Psychology and aging*, 26(3), 631–635. https://doi.org/10.1037/a0022852

Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Aging, Subjective Experience, and Cognitive Control: Dramatic False Remembering by Older Adults. *Journal of Experimental Psychology: General, 134(2)*, 131–148. https://doi.org/10.1037/0096-3445.134.2.131

JASP Team. (2020, May 19). *Bayesian inference in JASP: A new guide for students*. JASP – Free and User-Friendly Statistical Software. https://jasp-stats.org/2020/05/19/bayesian-inference-in-jasp-a-new-guide-for-students

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford University Press.

Karanian, J. M., Rabb, N., Wulff, A. N., Torrance, M. G., Thomas, A. K., & Race, E. (2020). Protecting memory from misinformation: Warnings modulate cortical reinstatement

during memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America,* 117(37), 22771–22779. https://doi.org/10.1073/pnas.2008595117

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kelley, C.M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language*, 48, 704-721.

Kim, E., Bayles, K.A., & Beeson, P.M. (2008). Instruction processing in young and older adults: Contributions of memory span. *Aphasiology, 22,* 753 - 762.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological review*, *103*(3), 490–517. https://doi.org/10.1037/0033-295x.103.3.490

Kouwenhoven, M., & Machado, L. (2023). Age differences in inhibitory and working memory functioning: limited evidence of system interactions. *Aging, Neuropsychology, and Cognition,* 31(3), 524–555. https://doi.org/10.1080/13825585.2023.2214348

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian data analysis for cognitive science: A practical course*. Cambridge University Press.

Light, L. L., Prull, M. W., La Voie, D. J., & Healy, M. R. (2000). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging (pp. 238–300).* Oxford University Press.

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

Loftus, E.F. (1981). Reconstructive Memory Processes in Eyewitness Testimony. In: Sales, B.D. (eds) *The Trial Process. Perspectives in Law & Psychology, vol 2. Springer, Boston, MA.* https://doi.org/10.1007/978-1-4684-3767-6_3

Loftus E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learning & memory (Cold Spring Harbor, N.Y.),* 12(4), 361–366. https://doi.org/10.1101/lm.94705

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

Macleod, M., & Saunders, J. (2017). Episodic Memory and Age-Related Deficits in Inhibitory Effectiveness. *Experimental Aging Research,* 43, 34 - 54. https://doi.org/10.1080/0361073X.2017.1258220.

Malmberg K. J. (2008). Recognition memory: a review of the critical findings and an integrated theory for relating them. *Cognitive psychology*, 57(4), 335–384. https://doi.org/10.1016/j.cogpsych.2008.02.004

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

McDonough, I., Bui, D., Friedman, M., & Castel, A. (2015). Retrieval monitoring is influenced by information value: the interplay between importance and confidence on false memory.. *Acta psychologica,* 161, 7-17 . https://doi.org/10.1016/j.actpsy.2015.07.017.

Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: how failing to retrieve practice items reduces overconfidence. *Consciousness and cognition, 29*, 131–140. https://doi.org/10.1016/j.concog.2014.08.008

Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology,* 17(1), 107–119. https://doi.org/10.1002/acp.857

Mueller-Johnson, K., & Ceci, S. J. (2004). Memory and Suggestibility in Older Adults: Live Event Participation and Repeated Interview. *Applied Cognitive Psychology, 18*(8), 1109–1127. https://doi.org/10.1002/acp.1078

Multhaup, K. S., De Leonardis, D. M., & Johnson, M. K. (1999). Source Memory and Eyewitness Suggestibility in Older Adults. *The Journal of General Psychology,* 126(1), 74–84. https://doi.org/10.1080/00221309909595352

Murphy, D. (2024). Responsible remembering: The role of metacognition, forgetting, attention, and retrieval in adaptive memory. *Psychonomic Bulletin & Review*, 32, 156 - 175. https://doi.org/10.3758/s13423-024-02554-9.

Murphy, D. H., & Castel, A. D. (2024). Serial and strategic memory processes in younger and older adults. *Aging, Neuropsychology, and Cognition,* 32, 207-236.

Murphy, D. H., Rhodes, M. G., & Castel, A. D. (2024). Age-related differences in metacognitive reactivity in younger and older adults. *Metacognition and Learning*, 19, 863-877.

Murphy, D. H., Schwartz, S. T., & Castel, A. D. (2022). Serial and strategic memory processes in goal-directed selective remembering. *Cognition*, 225, 105178.

Murray, B.D., Anderson, M.C., & Kensinger, E.A. (2015). Older Adults Can Suppress Unwanted Memories When Given an Appropriate Strategy. *Psychology and Aging,* 30, 9 - 25.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of experimental psychology. Learning, memory, and cognition*, *14*(4), 676–686. https://doi.org/10.1037//0278-7393.14.4.676

Nelson, T. O., & Narens, L. (1990). *Metamemory: A theoretical framework and new findings.* In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). Academic Press.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness Recall and Testimony. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 93–150). The Guilford Press.

Pansky, A., Goldsmith, M., Koriat, A., & Pearlman-Avnion, S. (2009). Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology, 21*, 303 - 329.

Payne, B. R., Jackson, J. J., Hill, P. L., Gao, X., Roberts, B. W., & Stine-Morrow, E. A. (2012). Memory self-efficacy predicts responsiveness to inductive reasoning training in older adults. *The journals of gerontology. Series B, Psychological sciences and social sciences,* 67(1), 27–35. https://doi.org/10.1093/geronb/gbr073

Prull, M. W., Dawes, L. L., Martin, A. M., 3rd, Rosenberg, H. F., & Light, L. L. (2006). Recollection and familiarity in recognition memory: adult age differences and neuropsychological test correlates. *Psychology and aging*, 21(1), 107–118. https://doi.org/10.1037/0882-7974.21.1.107

Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Rey-Mermet, A., & Gade, M. (2017). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review, 25,* 1695 - 1716. https://doi.org/10.3758/s13423-017-1384-7.

Roediger, H. L., 3rd, & Geraci, L. (2007). Aging and the misinformation effect: a neuropsychological analysis. *Journal of experimental psychology. Learning, memory, and cognition,* 33(2), 321–334. https://doi.org/10.1037/0278-7393.33.2.321

Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence–accuracy correlation. *Journal of Applied Psychology,* 81(5), 587–594. https://doi.org/10.1037/0021-9010.81.5.587

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

R. Rosenbaum, P. (2020). The Power of a Sensitivity Analysis and Its Limit. *Springer Series in Statistics.*

Salthouse T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society :* JINS, 16(5), 754–760. https://doi.org/10.1017/S1355617710000706

Tendeiro, J. N., & Kiers, H. A. L. (2019). A review of issues about null hypothesis Bayesian testing. Psychological Methods, 24(6), 774–795. https://doi.org/10.1037/met0000221

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

Tiego, J., Testa, R., Bellgrove, M., Pantelis, C., & Whittle, S. (2018). A Hierarchical Model of Inhibitory Control. *Frontiers in Psychology,* 9. https://doi.org/10.3389/fpsyg.2018.01339.

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., Dablander, F., Gronau, Q. F., Kucharský, Š., Gupta, A. R. K. N., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E.-J. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique, 120*(1), 73–96. https://doi.org/10.3917/anpsy1.201.0073

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic bulletin & review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

Vredeveldt, A., & Hollins, T. J. (2015). Metacognition moderates the effects of distraction on cognition. *Frontiers in psychology, 6,* 106. https://doi.org/10.3389/fpsyg.2015.00106

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E. J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic bulletin & review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wahlheim, C. N., Ball, B. H., & Richmond, L. L. (2017). Adult age differences in production and monitoring in dual-list free recall. *Psychology and aging*, 32(4), 338–353. https://doi.org/10.1037/pag0000165

Wells, G. L., & Quinlivan, D. S. (2009). Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. *Law and human behavior*, *33*(1), 1–24. https://doi.org/10.1007/s10979-008-9130-3

Wylie, L. E., Patihis, L., McCuller, L. L., Davis, D., Brank, E. M., Loftus, E. F., & Bornstein, B. H. (2014). Misinformation effect in older versus younger adults: A meta-analysis and review. In M. P. Toglia, D. F. Ross, J. Pozzulo, & E. Pica (Eds.), *The elderly eyewitness in court* (pp. 38–66). Psychology Press.

Yonelinas A. P. (2001). Consciousness, control, and confidence: the 3 Cs of recognition memory. *Journal of experimental psychology. General*, *130*(3), 361–379. https://doi.org/10.1037//0096-3445.130.3.361

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, 46(3), 441-517.