6-2022

# Evaluating the Effects of Initial Testing on Misinformation Suggestibility for Eyewitnessed Videos

Wryleigh Shearin-Anderson
*The University of Souther Mississippi*

# Evaluating the Effects of Initial Testing on Misinformation Suggestibility for Eyewitnessed Videos

by

Wryleigh Shearin-Anderson

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of Honors Requirements

May 2022

Approved by:

Mark Huff, Ph.D., Thesis Advisor,
School of Psychology

Sara Jordan, Ph.D., Director,
School of Psychology

Sabine Heinhorst, Ph.D., Dean
Honors College

# ABSTRACT

Exposure to misleading details following an eyewitnessed event often leads to memory errors for these misleading details—a pattern termed the misinformation effect. A recent debate is whether completing a memory test after a witnessed event, but before exposure to misleading details, can reduce subsequent misinformation (a protective effect of testing; PET) or increase subsequent misinformation (retrieval-enhanced suggestibility; RES). We further evaluated the initial testing effects using witnessed videos (vs. static images) which often yield a RES pattern and using household scenes which often yield a PET pattern. Following study of four household videos (e.g., bathroom, bedroom, etc.) that depicted an actor interacting with a set of objects, participants either completed an initial recall test or a filler task (no test control), followed by exposure to misinformation in the form of false objects and a final recall and source-monitoring test. Experiment 1 had participants complete the misinformation/final test phases during the same experimental session, whereas Experiment 2 delayed the misinformation/final test phases by 48 hours. In both experiments, initial testing improved correct memory for presented objects in the videos but had no effect on misinformation in either recall or source tests. Our results, therefore, indicate that while testing can benefit correct memory, it does not produce a memory cost (i.e., RES) to misinformation.

***Keywords:*** Misinformation; Retrieval Enhanced Susceptibility; Protective Effect of Testing; Videos; Delay; BFI

# DEDICATION

This thesis is dedicated to my supportive friends and family who spent hours of their time listening to me ramble on about my topic. I am beyond grateful for their love and encouragement throughout my time at The University of Southern Mississippi.

And to my cats, Poppy, Hawthorne, Sabre, and Jersey, for which I received a college education so I can give them a better life.

# ACKNOWLEDGMENTS

I would like to acknowledge my thesis advisor, Dr. Mark Huff, for the guidance he has given throughout the process of writing this thesis. I would like to thank Nick Maxwell for the time and effort he put into creating programs to run the experiments, compiling the data that was gathered, and for his feedback. Finally, I would like to thank the Drapeau Center for Undergraduate Research for the SPUR Grant that funded my research and brought my thesis to fruition.

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ANOVA | Analysis of Variance |
| PET | Protective Effect of Testing |
| RES | Retrieval Enhanced Suggestibility |

# CHAPTER I: Introduction

Memory researchers in both basic and applied settings have been interested in the malleability of human memory at least back to the 1970s. One contributing factor is the presence of misleading or false details about a previously experienced event which can be incorporated into subsequent retrievals—a pattern termed the *misinformation effect* (Loftus et al., 1978; Zaragoza et al., 2007). Misinformation is particularly problematic in an eyewitness context in which witnesses, who are often queried (i.e., tested) for details shortly after an event has occurred, must later recount a witnessed event after a delay such as providing testimony in court. If eyewitnesses are susceptible to misinformation, these errors could have drastic consequences. My thesis will further evaluate misinformation effects on memory accuracy by examining whether completing an initial memory test immediately following a witnessed video affects the susceptibility for interfering misinformation details and whether misinformation is affected by the delay of the initial test.

Misinformation paradigms generally follow a three-stage procedure in which participants are presented with an original event, are exposed to misleading details about that event, and are then tested for the original event. On the final test, participants are more likely to report or endorse misleading false details relative to a group or condition that is not presented with misleading details. The misinformation effect is reliable and is found when misinformation is embedded both within a series of questions (Saunders & Jess, 2010) and in a narrative form (Takarangi et al.,2006), occurs when the misinformation originates from different sources aside from the experimenter including highly credible sources (Numbers et al., 2014; Skagerberg & Wright, 2009), and social others, both real (Roediger et al., 2001) and perceived (Meade & Roediger, 2002; Huff et al., 2013). The misinformation effect also persists after a delay following exposure to misleading details (Frost et al., 2002; Schwartz & Wright, 2012), and is resistant to warnings which can successfully reduce, but do not eliminate, the effect (Chambers & Zaragoza, 2001; Eakin et al., 2003; Echterhoff et al., 2005). Indeed, nearly 50 years of research has consistently confirmed that exposure to misleading details is costly to subsequent memory accuracy and that this pattern generalizes across many different contexts in both basic and applied settings.

Misinformation costs to memory have also encouraged research interest in methods that may inoculate memory for the original event when exposed to false details. For instance, a meta-analysis (Blank & Launay, 2014) has revealed that post-warnings, or warnings that occur after misinformation is encountered but before a final test, can reduce misinformation by more than 40% relative to unwarned conditions. Despite these benefits, however, warnings may not be pragmatic to implement given it may be unknown whether participants were exposed to incorrect information or not. Further, warnings may also discourage participants from reporting details including those that may be correct due to concerns that they may report incorrect information. Separately, minimizing forgetting of the original event such as minimizing the delay between the original event and misinformation exposure has been shown to reduce the misinformation effect (Belli et al., 1992), as has ensuring participants have full attention available to encode a witnessed event (Lane, 2006). Moreover, misinformation items that lend themselves to successful memory monitoring are also less likely to be falsely remembered than those that are not. For instance, misinformation effects are reduced for misleading details that are central to an original event (Heath & Erickson, 1998; Paz-Alonso & Goodman, 2008; but see Dalton & Daneman, 2006, for a contrasting pattern), and when misleading details directly contradict a specific detail from the original event and are more likely to be detected as false (Huff & Umanath, 2018). However, like warnings, eyewitness situations do not place the above conditions under the control of the witness and are therefore unlikely to be applied in practice.

An ideal method for reducing misinformation would therefore be one that could be applied following the witnessed event but prior to exposure, or potential exposure, of misleading details. One such method that has received recent attention is that of initial testing. Testing has been shown to be highly effective at enhancing retention for studied materials over restudy (Roediger & Karpicke, 2006)—a pattern termed the retrieval-practice effect. Retrieval practice has been shown to facilitate the memory strength of information that is initially retrieved (Kornell et al., 2011) which can slow its subsequent forgetting. Several mechanisms have been used to account for testing benefits including enhanced organization via relational processing (Congleton & Rajaram, 2012), the implicit generation of mediators which could later be used as retrieval cues (Pyc & Rawson, 2010), and improved memory strength of information that is initially retrieved (Kornell et al., 2011). Regardless of the mechanisms, however, memory benefits following testing are robust and have been found across a variety of materials, delays, and test types (see Rowland, 2014 for a review and meta-analysis).

Paradoxically to the benefits of testing on facilitating correct memory, some misinformation studies have reported that initial testing can reduce overall memory accuracy by increasing suggestibility. This pattern was initially deemed a reversed testing effect but has since been termed *retrieval-enhanced suggestibility* (RES; Chan et al., 2009; Chan et al., 2017). In their initial study, Chan et al. (2009) had participants view an episode of a television series which was either followed by an initial cued-recall test for details within the episode, or no initial test. Following a post-event audio narrative that contained details that were both misleading and consistent with details from the television episode, participants who completed an initial test were *more* likely to false recall misleading details than participants who did not complete an initial test. Subsequent studies have reported the RES pattern using a variety of methodological conditions including a week delay between study and misinformation phases (Chan & Langley, 2011), using different types of videos that differ in duration (40 min vs. 8 min; Whilford et al., 2014), and when different initial test types are used including the cognitive interview which is often promoted by researchers as a beneficial method for probing memories for eyewitnesses (LaPaglia et al., 2014). RES may however be less likely to occur when misinformation is presented via misleading questions rather than embedded within a post-event narrative (LaPaglia & Chan, 2014). Regardless, RES produces a memory cost that can occur despite well-established benefits of retrieval practice.

RES patterns are typically interpreted via two mechanisms which are not mutually exclusive. First, the attentional account indicates that RES occurs because initial-test participants approach the misinformation phase as containing corrective feedback and are more likely to attend to misleading details. Consistent with the possibility, RES patterns are eliminated (but not reversed) when participants are warned that a narrative may contain false details (Thomas et al., 2010). Warnings may reduce the likelihood that participants approach the narrative as containing corrective feedback and reduce attention allocated to misleading items. Additionally, Gordon and Thomas (2014; see too Gordon et al., 2020) reported that RES patterns were associated with longer reading times for narrative statements with misleading details, suggesting greater attention is allocated to false details. Separately, a test-potentiated learning account (Cho et al., 2017; Gordon & Thomas, 2014; Pastotter & Bauml, 2014) posits that testing facilitates the encoding of new information, which includes the encoding of subsequent misleading details.

Despite studies indicating that initial testing can be harmful to memory accuracy in misinformation paradigms, a growing set of studies indicates that initial testing can reduce misinformation, a pattern termed a *protective effect of testing* (PET; Huff et al., 2013; 2016; Pereverseff et al., 2020). A PET effect was initially reported by Loftus (1978) in which participants viewed a car accident scene and some of the participants were questioned about the car's color before being provided misinformation about the color. During a final test, participants who were initially tested were more accurate in retrieving the original color than those who were not, indicating a benefit of initial testing. Loftus (1979) similarly proposed a "freeze effect" of initial testing, referring to a set of pilot data in which completing an initial recall test for an initial event would make the memory for this initial event more resistant to misinformation (see too, Howe, 1970, for a "freezing" pattern in a repeated reproduction paradigm). More recently, the cognitive interview, which was found to produce RES (LaPaglia et al., 2014), has also been found to reduce later misinformation when completed before (vs. after) exposure to misleading details (Memon et al., 2010). Other studies (Gabbert et al., 2012; Pansky & Tenenboim, 2011) have shown this same reduction in misinformation following an initial test. Taken together, initial testing can also reduce subsequent misinformation and improve memory accuracy, consistent with the benefits of retrieval practice on correct memory.

In the present study, my thesis will further examine the effects of initial testing on subsequent misinformation using the *social contagion of memory paradigm* (Roediger et al., 2001). The social contagion paradigm was initially adapted by Huff et al. (2013) to evaluate initial testing effects on subsequent misinformation that was presented from a social source. In their study, participants studied six static images of household scenes (bathroom, kitchen, etc.) which contained a variety of objects. Following study, some participants completed an initial free-recall test for objects in each scene, while other participants completed a time-matched filler task (i.e., the no-test group). Both groups then viewed a set of fictitious recall tests that were completed by “other participants from a prior study” that contained false supplemental details (i.e., contagion items) that were schematically consistent with the studied scenes. Exposure to these fake tests was followed by a final free-recall test and a source-monitoring test (cf. Johnson et al., 1993) in which participants were provided with a list of objects including the contagion items and were tasked with specifying whether the objects were presented in the original scenes, recalled by the participants, both, or not presented anywhere in the experiment. On the final free-recall test, a retrieval-practice effect was found for correct recall, but there was no difference in misinformation recall between the initial-test and no-initial-test groups. On the source-monitoring test, however, a PET pattern was found in which initial testing reduced the likelihood that participants would falsely ascribe the source of contagion items to original scenes.

Subsequent experiments using the household images in the social contagion paradigm indicated that the PET effect on the source test is found when participants take one or two initial tests and are exposed to misinformation either immediately after the initial test phase or a 48-hour delay (Huff et al., 2016). A PET effect was also found on a free-recall test, but only when the misinformation/final test phases were delayed. More recently, Pereverseff et al. (2020), using the same social contagion paradigm found a source-test PET effect when misleading items were embedded in both misleading questions and narrative contexts, indicating that the mode of misinformation delivery produces a consistent pattern.

Because initial testing can either increase or decrease suggestibility to misinformation, researchers have attempted to reconcile these differences by evaluating whether RES versus PET effects are unique to one type of methodology. One potential factor is the type of initial and final memory tests that have been completed. Most RES studies have used cued-recall tests (Chan et al., 2017; Chan et al., 2009; Thomas et al., 2010), though RES has been found on several types of initial/final tests including free recall (Wilford et al., 2014), source-monitoring (Chan et al., 2012), and the cognitive interview (LaPaglia et al., 2014). Likewise, a PET pattern has been found using various tests including cued-recall (Pansky & Tenenboim, 2011), free recall (Gabbert et al., 2012; Huff et al., 2016; Loftus, 1979), recognition (Loftus, 1977), and source-monitoring (Huff et al., 2013; 2016; Pereverseff et al., 2020). RES and PET patterns, therefore, appear to be test-type invariant.

Differences in misinformation formatting have also been shown to be invariant to RES and PET patterns. For instance, although many RES studies have embedded misinformation within a narrative format (Chan & LaPaglia, 2011; Chan et al., 2009; 2017; Thomas et al., 2010; Wilford et al., 2014), studies that have found a PET pattern have utilized cued-recall questions (Gabbert et al., 2012; Lane et al., 2001; Pansky & Tenenboim, 2011) or used the social-contagion paradigm (Huff et al., 2013; 2016). Consistent with these differences, Chan and LaPaglia (2013; see too LaPaglia & Chan, 2019) reported that the presentation of misinformation in a written narrative induced a RES pattern, but the presentation of misinformation in misleading questions induced a PET pattern when participants initially viewed a video of a crime. However, Pereverseff et al. (2020) similarly compared misinformation that was embedded in narratives or misleading questions using the static images from the social-contagion paradigm. In two experiments a PET effect was found on a final source-monitoring test for both types of misinformation formats, indicating misinformation format may not be related to whether initial testing is costly or beneficial to memory accuracy.

An additional possibility, and one that is investigated in the present study, is whether the formatting of the original event—either as a presented video or as static images—may be related to RES and PET patterns. RES studies generally use video clips as witnessed events given their similarities to eyewitness scenarios (e.g., Chan et al., 2009; LaPaglia & Chan, 2019), while studies that have found PET effects typically use static images, such as slideshow pictures (Loftus 1978; Pansky & Tenenboim, 2011; Pereverseff et al., 2020). Although one study has reported a PET pattern when using a crime simulation video (Gabbert et al., 2012), it is possible that RES patterns may be more likely to occur with video materials and PET patterns with static images. The purpose of the present study is to further assess whether initial testing contributes to a PET or RES pattern but using video materials in the social contagion paradigm. Given Huff and colleagues (Huff et al., 2013; 2016; Pereverseff et al., 2020) have consistently found evidence for a PET pattern using static images in the social contagion paradigm, the present study maintains the social contagion paradigm, but tests for RES and PET patterns following a witnessed video.

Finally, previous research suggests that there may be a link between personality factors and cognitive processes that may affect how participants process suggested details in a misinformation paradigm. For example, Jackson and Balota (2012) reported that individuals high in Conscientiousness showed improved mind wandering and were more likely to remain on task in a sustained attention to response paradigm. This pattern suggests that individuals high in Conscientious may have better tuned attentional control and may be better able to focus on task-specific details. Given these attentional benefits, one possibility is that high Conscientious individuals may have better attention towards original event details and may be better able to detect misinformation when presented with post-event information. If so, high Conscientious individuals may be less likely to show suggestibility. More germane to misinformation effects however, Frost et al. (2013) reported a positive relationship between Agreeableness and false endorsement of misinformation details, particularly between Agreeableness questions pertaining to compliance. Additionally, Gudjonsson (1983) reported a positive association between Neuroticism and suggestibility. It is important to note that this latter relationship with Neuroticism was not found by Frost et al., which might be due to Gudjonsson using a different Neuroticism measure than what is commonly used in Big 5 studies, or that the relationship with Neuroticism is unreliable.

In the present study, the relationships between the Big 5 personality factors (Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness) are further examined to test for the reliability of the patterns above. Focus is given to Conscientiousness, Agreeableness, and Neuroticism, given these three factors have shown positive associations with misinformation suggestibility in previous studies. It is therefore predicted that these positive relationships will again be found but using a different set of misinformation materials based on the social contagion paradigm.

**The Present Study**

In two experiments, the effects of initial testing on subsequent misinformation suggestibility were evaluated using the social contagion paradigm. In Experiment 1, participants studied a set of videos which depicted a female actor walking through four household scenes (bathroom, bedroom, kitchen, and garage). These scenes contained a variety of schema-consistent objects of which the actor interacted with a subset. Following the presentation of the videos, half of the participants completed an initial scene-cued free recall test (i.e., the test group) while the other half completed a filler task (i.e., the no-test group). Following the initial test/filler phase, all participants were presented with a series of fake recall tests ostensibly completed by other participants which contained false items that were also schema consistent. After viewing the fake recall tests, all participants immediately completed a final scene-cued free recall test and a source recognition test where participants were required to specify the source for a set of recognition test items. Finally, participants completed the BFI-2 (Soto & John, 2017) personality inventory. The experiment was closely modeled after Huff et al.'s (2013; 2016) experiments with the exception that household videos were used rather than static household images. Overall, it was expected that if a RES pattern was due to the use of initial videos as witnessed events, then initial-test participants would be more suggestible to misleading items than no-test participants. However, if a PET pattern or a null effect of initial testing was found, then the modality of witnessed event would not be a contributing factor to when RES is found.  Finally, an additional goal of the study was to evaluate the relationship between BFI-2 (Soto & John, 2017) factors and measures of memory. As such, differences in false source responses and recall, and correct recall and source responses were correlated with the BFI-2 factors, and the relationship in both Experiments 1 and 2 were examined.

To provide a second test of whether initial testing following studied videos contributes to RES or PET patterns, Experiment 2 closely followed Experiment 1 but included a 48-hour delay between the initial test/filler task phase and the misinformation phase. Because delayed tests often strengthen correct memory benefits of retrieval practice (e.g., Roediger & Karpicke, 2006; Rowland, 2014), this delay was expected to increase the magnitude of any testing effect on subsequent misinformation. Additionally, the use of a delay was expected to improve the external validity of the experiments given re-testing of witnesses often occurs following a delay rather than within a single study session. It was therefore predicted that the results of Experiment 2 would follow Experiment 1 but that the testing effect on correct memory and misinformation would be exaggerated.

# CHAPTER II: Experiment 1

## Methods

### *Participants*

Eighty participants were recruited through Prolific (www.prolific.co), an online crowdsourced participant pool. Participants were compensated for the study at a rate of $9.50 per hour. Participants were randomly assigned to the Initial Test group or the No Test group ($n$ = 40 per group), however data from two participants were removed—one for extremely low performance (< 5% recall) and another due to a computer error which did not present one of the scenes, leaving 39 participants in each group. All reported normal or corrected-to-normal vision and were proficient English speakers.

### *Materials*

Study materials consisted of silent videos of household scenes and were taken from Gretz and Huff (2019). The four silent videos depicted a female actor walking through several common household rooms (bathroom, bedroom, kitchen, and garage) while interacting with a set of household objects. These videos were based on static household images used in the social contagion paradigm (Roediger, Meade & Bergman, 2001). The videos contained an average of 25.25 objects per video (range = 22–27), and the objects were normed to be typical objects for each scene using 18 undergraduate students (see Gretz & Huff, 2019, for norming information). Based on these norms, two items for each scene that were rated by participants as being highly typical for a given scene were excluded from each video. These items were designed as critical items and were later suggested during the misinformation phase. Critical items consisted of the following: toilet/lotion (bathroom), shoes/laptop (bedroom), blender/spatula (kitchen), and scissors/rope (garage).

Misinformation was suggested via false objects embedded within fake recall tests, which participants were informed had been completed by other participants during a "previous study." However, these tests were experimenter prepared and consisted of both items that were initially presented in the scenes and the non-presented critical items. The misinformation phase was modeled after Huff et al. (2013) who found elevated misinformation rates when embedding critical items in fake recall tests. Participants viewed the recall tests from four "other participants" that were presented electronically. The number of items per test ranged from 6 to 10 total items (8 average). Critical items were incorporated into recall tests from two of the scenes (misinformation scenes), while tests from the other two scenes only contained correct items (control scenes). The scenes that contained misinformation critical items (vs. control scenes) were counterbalanced across participants. Critical items were always presented in test positions 4 and 6, while the other test positions always contained correct items that were randomly sampled from the scenes. Participants viewed four recall tests (one from each scene) for each of the four fake participants for 16 total recall tests. For recall tests that contained critical items, these were recalled by all four fake participants (i.e., participants received four exposures to each critical item) to ensure a sizeable misinformation effect.

Participants completed a 42-item source monitoring test that contained items that may or may not have been presented during the study phase. Participants were instructed to report whether an item was touched by the actor, not touched, or not presented throughout the videos.

Finally, the 60-item BFI-2 (Soto & John, 2017) was used to gather personality data about individual differences among the participants. This measure evaluates openness, contentiousness, extraversion, agreeableness, and neuroticism by having participants provide self-ratings of descriptive statements (e.g., "Is complex, a deep thinker"). Participant responses were recorded using a 5-point Likert scale ranging from disagree strongly (1) to agree strongly (5) for each of the statements. Half of the items were reversed scored.

***Procedure***

Data collection was completed online using Collector (Garcia & Kornell, 2015). Figure 1 illustrates the study design. Following informed consent and collection of demographics information, participants were instructed that they would view a series of silent household videos that depicted an actor interacting with a variety of common household objects. Participants were further informed that their memory for the objects in the scene would be tested later and that they were to remember as many objects as possible, regardless of whether the actor interacted with the object or not. Each video was prefaced with a title screen denoting the location of the video (e.g., "Kitchen") followed by a screen that automatically played each video. Each video lasted approximately 45-50 s, and following the video's presentation, a "Next" button was activated which allowed participants to advance to the next title screen and video. Participants studied all videos in the same order (bedroom, bathroom, kitchen, and garage).

Upon completion of the videos, participants completed an initial two-minute filler task which consisted of naming as many U.S. states as they could from memory. After the filler task, participants in the No-Initial Test group completed four additional filler tasks (name as many celebrities as you can, name as many professional athletes, etc.). Each additional filler task lasted two minutes, resulting in an additional of eight minutes of filler task. Participants in the Initial-Test group were instructed to complete an initial free-recall test for each household scene. For this test, participants were presented with a screen with the name of the scene listed at the top and a dialogue box where they were to report as many objects as they could remember from the scene regardless of whether the object was touched by the actor or not. Participants were informed that spelling would not be counted against them and that they would have two minutes to recall for each video. Thus, time spent completing the free-recall tests in the Initial Test group was equivalent to the filler tasks in the No-Initial Test group. Following each recall test, participants advanced to the next test in which they were instructed to recall from a different video. This procedure was repeated until the Initial Test participants completed tests for all four videos.

After the No-Initial Test and the Initial-Test groups completed their filler tasks or recall tests, respectively, both groups began the misinformation phase. During this phase, participants were presented with misinformation in the form of four recall sheets which participants were informed had been completed by other participants during a previous study. However, these sheets were created by the researcher with false items embedded. To ensure participants attended to the recall items, participants provided pleasantness ratings for each item using a 5-point Likert scale ranging from very unpleasant (1) to extremely pleasant (5). Participants were not informed that these recall sheets contained misinformation.

Immediately following the misinformation phase, all participants completed a set of final free-recall tests which followed the same format as the tests completed by participants in the Initial-Test group. Following this test phase, participants then completed a source-monitoring recognition test. This test presented participants with a test object, and participants were asked to determine whether the object was "touched" or "not touched" by the actor in the original videos or whether the object was "not presented" in the videos. The source test consisted of 42 items of which 12 were objects that were correctly presented and touched by the actor (3 from each video), 12 correctly presented objects that were not touched (3 from each video) 4 falsely suggested critical items (2 from each misinformation scene), 4 critical items from the control scenes, and 10 household objects that were not presented in any of the videos or the fake recall tests which served as controls. Participants were required to provide a response for each test item. The source monitoring test was self-paced.

Following the source monitoring test, participants completed the 60-item Big Five Inventory-2 (Soto & John, 2017), which consisted of 12 items assessing each of the Big Five personality factors (neuroticism, extraversion, openness, agreeableness, and conscientiousness). Responses were made using a 5-item Likert scale (1 = strongly disagree; 5 = strongly agree) and, like the source test, participants were required to make a response for each item. Completion of this inventory was self-paced. Participants were then debriefed regarding the study including the exposure to false items that were embedded in the recall tests from other participants. A typical experimental session lasted approximately 45 min.

## Results

An alpha level of $p < .05$ was adopted for all statistical tests. Effect sizes were computed as partial-eta squared ($\eta_p^2$) for all Analyses of Variance (ANOVAs) and as Cohen's *d* for all *t*-tests.

### *Correct Recall*

Proportions of correct recall for the No-Initial test group and the Initial Test group for initial and final tests are reported in Table 1. A 2(Touched: Touched vs. Non-Touched) × 2(Test: Initial Test vs. Final Test) repeated-measures ANOVA was first used to examine recall differences within the initial-test group. Objects that were touched in the videos were recalled at a higher rate than objects that were not touched (.47 vs. .23; for the mean for touched and non-touched objects, respectively), $F(1, 38) = 118.75$, Mean Squared Error (*MSE*) = .02, $\eta_p^2 = .76$ and correct recall increased from the initial test to the final test (.30 vs. .40), $F(1, 38) = 78.57$, $MSE = .01$, $\eta_p^2 = .67$. An interaction was also found, $F(1, 38) = 4.85$, $MSE = .01$, $\eta_p^2 = .11$, which reflected a slightly larger increase on the final test recall for non-touched than touched objects.

To test for the presence of a retrieval-practice benefit, a 2(Touched: Touched vs. Non-touched) × 2(Test Group: Initial Test vs. No-Initial Test) mixed ANOVA was used to compare final recall performance between the two test groups. Touched objects were again recalled at a higher rate than non-touched objects (.46 vs. .25), $F(1, 76) = 210.08$, $MSE = .01$, $\eta_p^2 = .73$, and importantly, completing an initial test was found to improve correct recall overall relative to no-initial test (.40 vs. .32), $F(1, 76) = 4.94$, $MSE = .05$, $\eta_p^2 = .06$ – a benefit of initial testing (Huff et al., 2020; Roediger & Karpicke, 2006). The testing benefit was equivalent for touched and non-touched objects as the interaction was not reliable, $F < 1$. Final recall for the No-Initial Test group was also compared to the initial recall test in the Initial Test group. Only touched objects were recalled at a higher rate than non-touched objects (.43 vs. .20), $F(1, 76) = 215.42$, $MSE = .02$, $\eta_p^2 = .74$, and there were no differences between the two tests as both the main effect of test type and the interaction were not reliable, both $F$s < 2.46, $MSE$s < .05, $p$s > .12.

***False Recall***

Proportions of falsely suggested objects recalled on the final recall test are reported in Table 1. In this analysis, I compared proportions of false recall both in the videos in which misinformation was later suggested (misinformation present) and control videos in which misinformation was not encountered (control), the latter of which represents a baseline of misinformation recall in the absence of its suggestion. A reliable misinformation effect was found in recall in which suggested objects were falsely recalled at a higher rate than control objects (.31 vs. .04), $F(1, 76) = 64.05$, $MSE = .04$, $\eta_p^2 = .46$. The main effect of test group was marginal, $F(1, 76) = 3.21$, $MSE = .04$, $p = .08$, $\eta_p^2 = .04$, but importantly, a significant interaction was found, $F(1, 76) = 4.39$, $MSE = .04$, $\eta_p^2 = .06$. Specifically, initial testing was found to increase misinformation falsely recall relative to the No-Initial Test group, but only for scenes in which information was presented (.37 vs. .24), $t(76) = 2.05$, Standard Error of the Mean ($SEM$) = .06, $d = 0.46$, but not for control scenes (.03 vs. .04), $t < 1$. The increase in suggestibility for misinformation present videos following initial testing is consistent with a RES pattern.

***Source-Monitoring Recognition***

Table 2 reports the proportions of source attributions for misinformation items as a function of test group for each of the three source responses (whether the object was “touched,” “not touched,” or “not presented” in the videos). Total false source attributions were computed as misinformation items that were attributed as being either touched or not touched as responding to either of these source responses indicated that the participant was attributing the source of the object to the video.

To examine differences in total false recognition, a 2(Misinformation Type: Presented vs. Control) × 2(Test Group: Initial Test vs. No-Initial Test) mixed ANOVA was used. A reliable misinformation effect was also found on the source test as false object attributions to the videos were higher for suggested misinformation objects than control objects (.75 vs. .51), $F(1, 76) = 30.49$, $MSE = .09$, $\eta_p^2 = .29$. Unlike the pattern found in recall, however, testing had no effect on source misinformation, as neither the main effect of test group nor the interaction were reliable (both $F$s < 1). Correct source attributions for misinformation objects (computed as “not presented” attributions) were also analyzed. Correct attributions for misinformation objects did not differ between the test and no-test groups both for objects that were suggested (.26 vs. .22) and for misinformation control objects (.54 vs. .48), both $t$s < 1.

Correct source responses for objects that were presented in the videos were similarly analyzed (touched objects correctly attributed as “touched”; non-touched objects correctly attributed as “non-touched) as a function of test group. Proportions of correct source responses are reported in Table 3. An effect of object type was found, $F(1, 76) = 17.47$, $MSE = .05$, $\eta_p^2 = .19$, in which correct attributions were higher for touched objects than non-touched objects (.62 vs. .47). A marginal effect of test group was found, in which correct attributions were numerically greater in the Initial Test group than the No-Initial Test group (.57 vs. .51), $F(1, 76) = 3.32$, $MSE = .04$, $p = .07$, $\eta_p^2 = .04$. The interaction was not significant, $F(1, 76) = 1.83$, $MSE = .05$, $p = .18$. Thus, when taken together with false attributions for misinformation items initial testing generally had no effect on source recognition. Initial testing did not produce a RES pattern nor a PET pattern and testing only marginally (but not significantly) increased correct source attributions for objects that were presented in the videos.

***Correlations with the Big-5 Personality Inventory***

Individual differences in correct recall and source responses and false recall and source responses to misinformation items were then correlated with the BFI-2 (Soto & John, 2017). Mean BFI responses for each of the five factors as a function of test group are reported in Table 4 and a correlation matrix is reported in Table 5. Given the study goal of evaluating the relationships between the BFI-2 factors and memory measures, correlation analyses reported here focus on the relationships with the memory measures, but full correlations including those across the BFI-2 factors are reported in the correlation matrix. In the analyses, correlations are collapsed across initial test groups to maximize available statistical power to detect relationships. A separate analysis with test group as a factor yielded no interactions ($F$s < 1), providing statistical justification for a pooled analysis.

Starting with correlations between misinformation recall and false source attributions of misinformation items, only a marginal positive relationship emerged between Agreeableness and false recall of misinformation items ($r = .21$, $p = .06$), with all other correlations being non-significant ($r$s < .16, $p$s > .16). Although this pattern is not reliable, it may suggest that individuals who are more agreeable may also be more suggestible when presented with false information from a perceived social source.

Moreover, this marginal pattern was only found on the free recall test (agreeableness was not marginally related to misinformation source attributions; $r = .03$, $p = .81$) which may indicate a potential test-type difference from when participants are freely able to retrieve from memory versus test types that require specification of contextual details.

Turning to correct memory, no significant correlations were found between any of the BFI-2 factors and touched recall or touched source attributions ($r$s < -.17, $p$s > .13). However, for non-touched recall, a small positive relationship was found with Neuroticism ($r = .28$, $p = .01$) as was a small negative relationship with Conscientiousness ($r = -.24$, $p = .04$). A similar positive relationship with Neuroticism was found for correct non-touched source attributions ($r = .25$, $p = .03$), however the relationship with Conscientiousness was not reliable ($r = -.18$, $p = .11$). This consistent pattern indicates that individuals who were generally more worried and anxious were more attuned to non-touched objects as they were better able to recall them and correctly retrieve their context on the source test. An additional relationship that emerged was a negative correlation between Extraversion and correct non-touched source retrievals ($r = -.25$, $p = .03$), which indicates that more introverted individuals are more likely to remember the source of non-touched items. This pattern is speculated on further in the Discussion, but these patterns may suggest some possible relationships between personality factors and correct memory retrievals.

**Discussion**

The purpose of Experiment 1 was to evaluate the effects of initial testing on subsequent misinformation using video materials in the social contagion paradigm. It was predicted that if initial testing increased misinformation (i.e., a RES pattern), this would provide evidence for initial video materials as contributing to greater misinformation suggestibility following testing. If, however, initial testing decreased misinformation (i.e., a PET pattern), this would provide evidence that the modality of study materials was not related to testing effects on misinformation as a PET pattern has been found reliably with static images (Huff et al., 2013; 2016; Pereverseff et al., 2020). Overall, the results of Experiment 1 were mixed. Although initial testing produced a RES pattern on the final free-recall test, this pattern did not extend to the source-monitoring test, as no differences were found between groups regarding false source attributions for misinformation objects.

While testing effects on misinformation were inconsistent across tests, initial testing generally benefitted correct memory for objects in the video. Initial test participants were more likely to correctly recall objects presented in the videos (i.e., retrieval practice; Roediger & Karpicke, 2006), and they were marginally more likely to correctly attribute touched and non-touched objects as being presented in the original videos. These two patterns indicate that initial testing may lead to memory costs regarding misinformation but may also produce memory benefits.

Correlations with the Big 5 Inventory produced different relationships between misinformation and correct memory. Starting with misinformation, patterns were weak and largely unreliable with only a positive trend found between Agreeableness and misinformation recall. This pattern is consistent with Frost et al. (2013) who also reported a positive relationship with Agreeableness, however this pattern was not found with suggestibility in the source test. Thus, the relationship between Agreeableness and misinformation may not occur when source details are queried. These patterns however were inconsistent with the prediction that both Conscientious and Neuroticism would be positively related to misinformation as neither of these relationships were reliable. For correct memory, Neuroticism was positively correlated to both non-touched recall and non-touched source attributions, indicating that more neurotic individuals may be more likely to successfully encode and retrieve objects that were less focal in the videos given the actor did not interact with them. Conscientiousness was negatively related to recall of non-touched objects, but this pattern did not persist into the source test, suggestion that Conscientiousness was less consistently associated with correct memory than Neuroticism.

Given the mixed evidence for initial testing effects on misinformation, Experiment 2 was designed to exaggerate testing effects on subsequent recall and source recognition. Previous research has shown that testing effects on correct memory are often more robust when a final test is delayed relative to being completed immediately (e.g., Karpicke & Roediger, 2006; Rowland, 2014, etc.). If testing is indeed more impactful following a delay, then misinformation may similarly be more sensitive to testing delays. Consistent with this possibility, Huff et al. (2016) reported that a reliable PET effect emerged on free recall when a 48-hour delay occurred between the initial-test phase and the misinformation and final test phases, but not when the misinformation/final test phases occurred immediately. Moreover, final test delays are likely common in eyewitness situations in which witness statements may not be gathered immediately following events. The effects of delays on testing are therefore examined in Experiment 2 in which a 48-hour delay was similarly inserted between the misinformation phase and the final recall/source-monitoring test phase.

# CHAPTER III: Experiment 2

Experiment 2 further examined the effects of initial testing on subsequent misinformation but inserted a 48-hour delay between the initial test phase and the misinformation and final test phases. Because testing often slows forgetting, it was expected that initial testing would be more impactful on correct memory and misinformation on a final test following a delay. The predictions aligned with Experiment 1: The emergence of an RES pattern would suggest that the video presentation of the studied event contributes to increased susceptibility to misinformation, whereas a PET pattern would suggest that video materials are not a contributing factor. Like Experiment 1, participants again completed the BFI-2. Based on the results of Experiment 1 (and Frost et al., 2013), it was expected that a positive relationship would emerge between Agreeableness and misinformation suggestibility, though Conscientiousness and Neuroticism relationships were also examined given the previous patterns in the literature (cf. Gudjonsson, 1983; Jackson & Balota, 2012).

## Methods

### *Participants*

One hundred and two participants were recruited through Prolific (www.prolific.co), and The University of Southern Mississippi's undergraduate research pool. Participants were compensated for the study at a rate of $9.50 per hour or with class credit. Participants were randomly assigned to two the Initial Test group ($n$ = 49) or the No Test group ($n$ = 53), All reported normal or corrected-to-normal vision and were proficient English speakers.

### *Materials*

The materials were the same as in Experiment 1.

### *Procedure*

The procedure was the same as Experiment 1, with one exception. There was an addition of a 48-hour delay between the initial recall and the misinformation phase for all participants.

## Results

### *Correct Recall*

Proportions of correct recall for the No-Initial Test group and the Initial-Test group for the initial and final tests in Experiment 2 are reported in Table 1. A 2(Touched) × 2(Test) repeated-measures ANOVA was used to analyze recall differences between the initial test and the final test within the Initial Test group. Recall rates for touched objects in the videos were higher than non-touched objects (.50 vs. .28; for the mean for touched and non-touched objects, respectively), $F(1, 48) = 202.90$, $MSE = .01$, $\eta_p^2 = .81$, and correct recall increased from the initial test to the final test (.37 vs. .41), $F(1, 48) = 10.04$, $MSE = .01$, $\eta_p^2 = .17$. An interaction was also found, $F(1, 48) = 39.00$, $MSE = .01$, $\eta_p^2 = .45$. Follow-up tests indicated that the increased recall on the final test was only found on non-touched objects (.24 vs. .31, for initial and final tests, respectively), $t(48) = 48) = 6.41$, $SEM = .01$, $d = 0.62$, but not for touched objects (.50 vs. .50), $t < 1$. The increase between initial and final tests despite a delay is likely due to participants being exposed to correctly studied objects during the misinformation phase.

To evaluate the presence of a retrieval-practice benefit, a 2(Touched) × 2(Test Group) mixed ANOVA was used to compare final recall performance between the two test groups. Touched objects were again recalled at a higher rate than non-touched objects

(.45 vs. .29) $F(1, 100) = 239.26$, $MSE = .01$, $\eta_p^2 = .79$. Similar to Experiment 1, initial testing led to higher correct recall than no initial testing (.41 vs. .33), $F(1, 100) = 7.03$, $MSE = .04$, $\eta_p^2 = .07$. A marginal interaction was also found, $F(1, 100) = 3.14$, $MSE = .01$, $p = .08$, $\eta_p^2 = .03$, which reflected a slightly larger testing benefit for touched objects (.41 vs. .50), $t(100) = 2.75$, $SEM = .03$, $d = 0.55$, than non-touched objects (.26 vs. .31), $t(100) = 2.15$, $SEM$ .02, $d = 0.43$. Finally, recall for the No-Initial-Test group was compared to the initial recall test in the Initial-Test group. Touched objects were again recalled at a higher rate than non-touched objects (.45 vs. .25), $F(1, 100) = 269.56$, $MSE = .01$, $\eta_p^2 = .73$. The main effect of test type was not reliable, $F(1, 100) = 1.76$, $MSE = .04$, $p = .19$, but a significant interaction was found, $F(1, 100) = 21.49$, $MSE = .01$, $\eta_p^2 = .18$. This interaction reflected higher recall for touched objects during the initial test in the Initial-Test group than the recall test in the No-Initial-Test group (.50 vs. .41), $t(100) = 2.72$, $SEM = .03$, $d = 0.54$, but no difference between the tests for non-touched objects (.24 vs. .26), $t < 1$.

### ***False Recall***

Table 1 also reports proportions of falsely suggested objects recalled on the final test. Again, a reliable misinformation effect was found, with suggested objects falsely recalled at a higher rate than non-suggested control objects (.36 vs. .07), $F(1, 100) = 66.49$, MSE = .09, $\eta_p^2 = .40$. However, unlike Experiment 1, initial testing had no effect on misinformation rates as neither the main effect nor the interaction were reliable, both $F$s < 1. Thus, following a delay, neither a RES nor a PET pattern were in evidence.

### ***Source-Monitoring Recognition***

Proportions of source attribution for misinformation objects as a function of test group are reported in Table 2. Total scores for false source attributions were calculated as in Experiment 1. A reliable misinformation effect was again found on the source test with a higher rate of suggested objects falsely attributed to the videos than control objects (.89 vs. .51), $F(1, 100) = 126.68$, $MSE = .06$, $\eta_p^2 = .56$. Replicating Experiment 1, testing produced no effect on suggestibility as both the main effect of testing and the interaction were unreliable, both *F*s < 1. Correct source attributions of suggested objects (i.e., “not presented” source responses) were compared between test groups. Correct attributions were equivalent between the Initial Test and No-Initial test groups, both for objects that were suggested (.14 vs. .09), $t(100) = 1.46$, $SEM = .04$, $p = .15$, and for misinformation control objects (.48 vs. .50), $t < 1$.

Finally, correct source attributions for objects presented in the videos were analyzed between test groups (proportions reported in Table 3). Correct source attributions did not differ between touched and non-touched objects (.54 vs. .50), $F(1, 100) = 2.08$, $MSE = .04$, $p = .15$, however, initial testing improved correct source attributions overall relative to the No-Initial-Test group (.55 vs. .50), $F(1, 100) = 5.90$, $MSE = .03$, $\eta_p^2 = .06$, indicating a retrieval-practice benefit on the source-recognition test. The interaction was not significant, $F(1, 100) = 1.51$, $MSE = .04$, $p = .22$. Thus, completing an initial test appeared to improve overall final test accuracy relative to no initial testing by improving both correct recall and source recognition while producing no effect on misinformation (i.e., no RES pattern).

***Correlations with the Big-5 Personality Inventory***

Correlations between memory responses and factors in the BFI-2 (Soto & John, 2017) were again computed. Table 4 reports the mean BFI-2 responses for the five factors as a function of test group. Table 6 reports the correlation matrix representing the relationship between memory measures and the BFI-2 factors.

Like Experiment 1, to maximize available power to detect relationships between personality measures and memory responses, correlations were collapsed across the two test groups (though again, test group did not moderate these relationships, $F$s $< 1$). Starting with misinformation recall and source misattributions to suggested objects, there were no significant correlations with any of the personality factors (largest $r = -.17$, $p = .10$). For correct memory, a significant relationship was found between Agreeableness and correct recall of touched objects ($r = .35$, $p < .001$), and correct recall of non-touched objects ($r = .22$, $p = .03$). A relationship was also found between Conscientiousness and correct recall of touched objects ($r = .21$, $p = .04$. However, for correct source attributions, only the relationship between Conscientiousness and correct source attributions of presented objects was found ($r = .22$, $p = .03$), suggesting that only Conscientiousness may be positively related to correct memory of touched objects consistently, following a delay.

**Discussion**

The goal of Experiment 2 was to evaluate the effects of initial testing on subsequent misinformation after a 48-hour delay using video materials. It was expected the misinformation pattern reported in Experiment 1 would be exaggerated following a delay due to participants' decreased memory for the initial event prior to receiving misinformation. As in Experiment 1, recall rates for touched objects were higher than non-touched objects, and additionally, correct recall increased from the initial test to the final test. Furthermore, a testing effect occurred in which initial testing led to higher correct recall than no initial testing with a greater benefit for touched objects over non-touched objects. Although a reliable misinformation effect was found, the delay used in Experiment 2 did not increase this effect as anticipated. Instead, there were no differences in misinformation rates between test groups. Regarding source attributions, testing produced no effect on suggestibility and instead produced a retrieval-practice benefit on the source-recognition test.

Turning to correlations with the personality measures, no relationships were found between any of the personality measures and misinformation suggestibility. This is counter to the positive relationship found in Experiment 1 and reported by Frost et al. (2013) between suggestibility and Agreeableness. Additionally, despite the initial predictions, neither Conscientiousness nor Neuroticism again were associated with suggestibility, indicating that the relationship between these personality factors and suggestibility are unreliable. Finally, for correct memory, Conscientiousness was positively related to touched object recall and source attributions, an observation in line with previous attentional findings by Jackson and Balota (2012).

# CHAPTER IV: General Discussion

Using videos based on the social contagion of memory paradigm (Roediger et al., 2001), the present study examined the effects of initial testing on subsequent misinformation. Following the presentation of the videos, half of participants were tested on objects presented in the videos while the other half were untested and completed a filler task. All participants were then presented with sets of false objects via an implied social source. Overall, completing an initial test generally improved both correct recall and correct source-monitoring for objects that were presented as participants were better able to attribute whether objects were touched or not by the actor in the videos. This retrieval-practice benefit (Roediger & Karpicke, 2006) was found both when final tests were completed within the same experimental session as the initial video presentation (Experiment 1) and when misinformation and final tests were delayed 48-hours (Experiment 2). Misinformation patterns, however, were largely test invariant. Although initial testing was found to increase suggestibility in recall in Experiment 1, this pattern was eliminated when participants specified source details in a subsequent recognition test. Furthermore, testing was found to have no effect on misinformation following the delay in Experiment 2.

A secondary purpose of this study was to evaluate potential personality moderators on correct memory and suggestibility using the Big 5 inventory. Based on previous literature, it was predicted the misinformation suggestibility would be positively related to Agreeableness (Frost et al., 2013), Neuroticism (Gudjonsson, 1983), and Conscientiousness, due to reported attentional benefits for individuals high in

Conscientiousness (Jackson & Balota, 2012). Although Experiment 1 yielded a positive relationship between Agreeableness and misinformation recall—a pattern consistent with Frost et al.—this pattern did not persist in Experiment 2 with a delay. Furthermore, no relationships were found between misinformation suggestibility and any other personality factor, including Neuroticism and Conscientiousness, suggesting that relationships between the Big 5 personality factors and misinformation suggestibility are either nonexistent or small and unreliable. Additionally, there were a few significant correlations that emerged with correct memory (e.g., Conscientiousness and Extraversion), but again, these were not found consistently across experiments.

The current experiments were designed to provide an additional evaluation of testing effects on subsequent misinformation given initial testing has been reported to both increase misinformation (i.e., RES; Chan et al., 2009; 2017; LaPaglia et al., 2014; LaPaglia & Chan, 2014) and decrease misinformation (i.e., PET; Memon et al., 2010; Pansky & Tenenboim, 2011; Gabbert, et al., 2012; Huff et al., 2013; 2016; Pereverseff et al., 2020). Although several methodological details have been evaluated (e.g., LaPaglia & Chan, 2014; Pereverseff et al., 2020), studies that have found a RES pattern generally use videos to depict an initial witnessed event, while studies that have reported a PET pattern (e.g., Huff et al., 2013; 2016) have used static images. The present study converted these static images into videos which were more likely to mimic the dynamic events that are depicted to real-world witnesses. Despite the potential modality patterns in the literature, initial testing generally had no effect on subsequent suggestibility, as a null testing pattern was the predominant finding in both experiments. Thus, initial testing following witnessing video materials does not appear to be a primary contributing factor to RES patterns, nor do video materials appear to contribute to a PET pattern.

One possible reason for the discrepant findings of RES and PET patterns may be due to the narrative component associated with the videos used in RES literature (e.g., an episode of *Flashpoint;* LaPaglia & Chan, 2019). Videos in these earlier studies contained a clear narrative component, unlike the videos used in the present study which simply depicted a person moving through various household scenes and interacting with objects with no clear narrative explaining the purpose of the movements or why some objects were touched relative to others. However, it is important to note that narrative features may not be consistently available in all eyewitnessed events. For example, for a crime that happens randomly, a witness may only view a snippet of the event and lack a narrative or details necessary to explain why the event occurred. Furthermore, common eyewitness events like car accidents or thefts are often random and unexpected, suggesting that story components may be exceptional features of witnessed events rather than standard. Therefore, if the utilization of a story narrative is responsible for RES patterns, then this could suggest that RES is a byproduct of the paradigm rather than a byproduct of an initial test.

Another explanation for this discrepancy in the occurrence of RES or PET may be due to the narrative of the misinformation instead of the initial event, as suggested by Chan et al. (2017). It has been found that when participants are presented with a cohesive narrative containing misinformation, a RES pattern has been found. However, when the narrative is made disjointed, this effect was eliminated (LaPaglia 2013; Experiment 3A). Building upon this, LaPaglia and Chan (2019) also compared the effects of initial testing on contextual information for questions and narratives. A RES pattern was found both for questions and when the contextual information was cohesive, and the information was reinstated during misinformation. The present study did not use a narrative to organize the presentation of objects in a post-event narrative and instead presented a set of seemingly random objects that were schematically consistent with the scenes including errors. While misinformation in the social contagion paradigm may be disjointed, it is important to note that Pereverseff et al. (2020) reported a PET pattern using the social contagion paradigm both when participants viewed misleading questions and a misleading narrative with false objects, suggesting that the cohesion within the post-event materials is not likely a contributing factor. It is possible, therefore, that some combination of these characteristics (i.e., initial video materials with a narrative; initial static images with misleading questions) may contribute to when RES/PET patterns emerge, and further, these combinations could be critical for when testing can help or hurt memory accuracy.

Although initial effects of testing were largely invariant on subsequent misinformation, it is important to emphasize that initial testing produced a net positive on memory accuracy. Indeed, the finding that initial testing can promote correct remembering of presented objects—both touched and non-touched—indicates that querying witnesses for their memories initially is beneficial. The present study adds to a large literature showing the benefits of retrieval practice while also highlighting how initial testing can promote correct memory for free-recall and source-monitoring in the face of suggested errors. Thus, testing appears to be an effective tool for promoting eyewitness memory.

## Limitations and Future Directions

Due to restrictions in place by the ongoing pandemic (Spring 2021 - Fall 2021), all participants completed the study online, not in a research lab facility. As a result, the lack of monitoring by an experimenter in a controlled environment may have resulted in participants paying less attention to the experiment or being vulnerable to distractions in their surroundings. Future studies should continue to assess factors driving RES and PET patterns using a controlled lab setting. Finally, future research may benefit from testing narrative versus non-narrative initial event study materials, as this may be a provide a key beginning in solving the RES and PET debate within the eyewitness misinformation literature while providing more insight into overall memory for witnessed events. Finally, future research may wish to continue the work of LaPaglia (2013). Because the present study presented misinformation in a disjointed manner via experimenter-made recall sheets, the addition of a narrative misinformation component may provide the potential to further examine these two patterns.

# CHAPTER V: Conclusion

Across two experiments, a testing benefit for correct memory occurred. Although in Experiment 1, a RES pattern emerged for scenes where misinformation was presented, this pattern was not in evidence in the source monitoring test or in Experiment 2 after a delay occurred. Instead, misinformation was found to be largely invariant to initial testing, supporting neither an RES pattern nor a PET pattern. Given the general benefit that testing had on correct recall and correct source attributions of objects presented in the scenes, testing was overall beneficial to memory accuracy and does not appear to be costly to eyewitness retrieval. Additionally, correlations with the Big 5 personality factors were assessed, however these factors were found to be inconsistent across experiments and with relatively small magnitudes of reliability. The Big 5 factors are therefore unlikely to be related to correct memory or misinformation effects following eyewitnessed events or when the relationships are small.

# APPENDIX A: FIGURES AND TABLES

*Figure 1*

*Depicts study design*

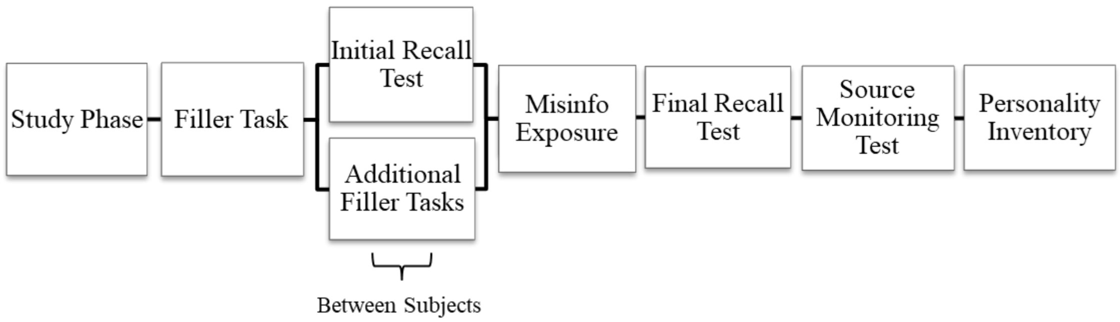

**Table 1**

*Mean (95% CI) Proportions of Correct Recall and False Recall of Misinformation Items and Extra-List Intrusions as a Function of Initial Test Group for Experiments 1 and 2.*

| | Experiment 1 (Immediate Test) | | Experiment 2 (Delayed Test) | |
|---|---|---|---|---|
| | No Initial Test Group | Initial Test Group | No Initial Test Group | Initial Test Group |
| *N* | 39 | 39 | 53 | 49 |
| Correct Recall | | | | |
| Touched | .42 (.06) | .51 (.06) | .43 (.03) | .51 (.07) |
| Not Touched | .22 (.04) | .29 (.05) | .25 (.04) | .29 (.05) |
| False Recall | | | | |
| Misinfo. | .24 (.06) | .37 (.10) | .43 (.10) | .40 (.11) |
| Misinfo. Control | .04 (.03) | .03 (.03) | .06 (.05) | .03 (.04) |

**Table 2**

*Mean (95% CI) Proportions of Source Attributions for Misinformation Items as a Function of Initial Test Group for Experiments 1 and 2.*

| | No Initial Test Group | | Initial Test Group | |
|---|---|---|---|---|
| | Misinfo. | Misinfo. Controls | Misinfo. | Misinfo. Controls |
| Experiment 1: Immediate Test | | | | |
| "Touched" | .36 (.10) | .20 (.07) | .33 (.09) | .19 (.07) |
| "Not Touched" | .42 (.10) | .31 (.10) | .40 (.09) | .26 (.09) |
| Total Errors | .78 (.10) | .51 (.12) | .73 (.10) | .45 (.11) |
| "Not Presented" | .22 (.10) | .48 (.12) | .26 (.10) | .54 (.11) |
| Experiment 2: Delayed Test | | | | |
| "Touched" | .41 (.08) | .15 (.06) | .35 (.08) | .15 (.06) |
| "Not Touched" | .51 (.07) | .37 (.08) | .52 (.08) | .35 (.07) |
| Total Errors | .92 (.05) | .53 (.09) | .87 (.06) | .50 (.09) |
| "Not Presented" | .08 (.05) | .48 (.09) | .14 (.06) | .50 (.09) |

**Table 3**

*Mean (95% CI) Proportions of Correct Source Attributions for Objects Correctly Presented in the Videos as a Function of Initial Test Group for Experiments 1 and 2.*

| | No Initial Test Group | Initial Test Group |
|---|---|---|
| Experiment 1: Immediate Test | | |
| “Touched” | .61 (.07) | .62 (.07) |
| “Not Touched” | .41 (.07) | .52 (.05) |
| “Not Presented” | .59 (.10) | .65 (.09) |
| Experiment 2: Delayed Test | | |
| “Touched” | .50 (.05) | .59 (.06) |
| “Not Touched” | .49 (.05) | .51 (.05) |
| “Not Presented” | .67 (.06) | .76 (.06) |

**Table 4**

*Mean BFI-2 responses as a function of test group*

| | No Initial Test Group | Initial Test Group |
|---|---|---|
| Experiment 1: Immediate Test | | |
| Openness | 3.65 | 3.72 |
| Conscientiousness | 3.15 | 3.68 |
| Extraversion | 3.94 | 3.03 |
| Agreeableness | 2.74 | 3.80 |
| Neuroticism | 3.82 | 3.11 |
| Experiment 2: Delayed Test | | |
| Openness | 3.72 | 3.72 |
| Conscientiousness | 3.56 | 3.68 |
| Extraversion | 2.95 | 3.03 |
| Agreeableness | 3.74 | 3.80 |
| Neuroticism | 3.17 | 3.11 |

**Table 5**

*Experiment 1 Correlations Between Variables*

| | Touched Recall | Non-Touched Recall | MI Presented | MI Control | MI Presented Source | MI Presented Correct | Correct Source Non-Touched | Correct Source Touched | N | E | O | A | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Touched Recall | - | .777** | .281* | -0.087 | -.251* | .248* | 0.216 | .313** | 0.1 | -0.17 | 0.136 | -0.02 | -0.139 |
| Non-Touched Recall | | - | .289* | -0.037 | -0.18 | 0.178 | .297** | 0.179 | .279* | -0.173 | 0.188 | -0.082 | -.240* |
| MI Presented | | | - | -0.088 | 0.136 | -0.161 | -0.014 | 0.112 | -0.172 | 0.115 | 0.16 | 0.21 | 0.113 |
| MI Control | | | | - | 0.195 | -0.19 | 0.082 | 0.07 | 0.064 | 0.04 | -0.126 | -0.023 | -0.027 |
| MI Presented Source | | | | | - | -.996** | 0.204 | 0.103 | 0.008 | 0.032 | -0.071 | 0.028 | 0.053 |
| MI Presented Correct | | | | | | - | -0.192 | -0.094 | -0.001 | -0.037 | 0.076 | -0.03 | -0.054 |
| Correct Source Non-Touched | | | | | | | - | -0.087 | .247* | -.248* | -0.127 | -0.163 | -0.181 |
| Correct Source Touched | | | | | | | | - | 0.091 | -0.097 | 0.023 | -0.052 | -0.041 |
| N | | | | | | | | | - | -.594** | -0.149 | -.587** | -.642** |
| E | | | | | | | | | | - | .282* | .392** | .434** |
| O | | | | | | | | | | | - | .428** | 0.162 |
| A | | | | | | | | | | | | - | .543** |
| C | | | | | | | | | | | | | - |

*Notes. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed). Touched Recall = correct recall for touched items on a final test, Non-Touched Recall = correct recall for non-touched items on a final test, MI Presented = Misinformation that was presented to participants that were reported on a final test, MI Control = Misinformation that was not presented to participants that were reported on a final test, MI Presented Source = Misinformation that was falsely attributed to the videos on the source monitoring test, MI Presented Correct = Misinformation that was correctly attributed to "not presented," Correct Source Non-Touched =Items that were correctly labeled as "not touched" on the source monitoring test, Correct Source Touched = Items that were correctly labeled as "touched" on the source monitoring test, N = Neuroticism, E =Extraversion, O = Openness, A = Agreeableness, C = Contentiousness*

**Table 6**

*Experiment 2 Correlations Between Variables*

| | Touched Recall | Non-Touched Recall | MI Presented | MI Control | MI Presented Source | MI Presented Correct | Correct Source NonTouched | Correct Source Touched | N | E | O | A | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Touched Recall | - | .774** | -.036 | .118 | -.182 | .183 | .082 | .516** | -.165 | .059 | .123 | 0.345** | .207* |
| NonTouched Recall | | - | -.023 | 0.284** | -.042 | .048 | .092 | 0.344** | -.136 | -.019 | .168 | .222* | -.041 |
| MI Presented | | | - | -.092 | .395** | -.391** | -.052 | .085 | .123 | -.166 | .062 | .100 | .010 |
| MI Control | | | | - | .163 | -.157 | .089 | -.034 | .047 | -.089 | -.035 | .031 | -.194 |
| MI Presented Source | | | | | - | -.992** | -.025 | .034 | .048 | .041 | .067 | -.126 | -.012 |
| MI Presented Correct | | | | | | - | .043 | -.037 | -.040 | -.043 | -.086 | .122 | .002 |
| Correct Source NonTouched | | | | | | | - | -224* | .086 | .077 | -.023 | -.081 | -.009 |
| Correct Source Touched | | | | | | | | - | -.037 | .113 | .013 | .183 | .218* |
| N | | | | | | | | | - | -.321** | .010 | -.189 | -.348** |
| E | | | | | | | | | | - | .131 | .004 | .373** |
| O | | | | | | | | | | | - | .121 | .142 |
| A | | | | | | | | | | | | - | .368** |
| C | | | | | | | | | | | | | - |

*Notes. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed). Touched Recall = correct recall for touched items on a final test, Non-Touched Recall = correct recall for non-touched items on a final test, MI Presented = Misinformation that was presented to participants that were reported on a final test, MI Control = Misinformation that was not presented to participants that were reported on a final test, MI Presented Source = Misinformation that was falsely attributed to the videos on the source monitoring test, MI Presented Correct = Misinformation that was correctly attributed to "not presented," Correct Source Non-Touched =Items that were correctly labeled as "not touched" on the source monitoring test, Correct Source Touched = Items that were correctly labeled as "touched" on the source monitoring test, N = Neuroticism, E =Extraversion, O = Openness, A = Agreeableness, C = Contentiousness*

# APPENDIX B: IRB APPROVAL LETTER

Office *of*
Research Integrity

THE UNIVERSITY OF SOUTHERN MISSISSIPPI.

118 COLLEGE DRIVE #5125 • HATTIESBURG, MS | 601.266.6576 | USM.EDU/ORI

**NOTICE OF INSTITUTIONAL REVIEW BOARD ACTION**

The project below has been reviewed by The University of Southern Mississippi Institutional Review Board in accordance with Federal Drug Administration regulations (21 CFR 26, 111), Department of Health and Human Services regulations (45 CFR Part 46), and University Policy to ensure:

- The risks to subjects are minimized and reasonable in relation to the anticipated benefits.
- The selection of subjects is equitable.
- Informed consent is adequate and appropriately documented.
- Where appropriate, the research plan makes adequate provisions for monitoring the data collected to ensure the safety of the subjects.
- Where appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of all data.
- Appropriate additional safeguards have been included to protect vulnerable subjects.
- Any unanticipated, serious, or continuing problems encountered involving risks to subjects must be reported immediately. Problems should be reported to ORI via the Incident template on Cayuse IRB.
- The period of approval is twelve months. An application for renewal must be submitted for projects exceeding twelve months.
- Face-to-Face data collection may not commence without prior approval from the Vice President for Research's Office.

PROTOCOL NUMBER: IRB-21-37
PROJECT TITLE: The effects of initial testing on misinformation for video study materials
SCHOOL/PROGRAM: Psychology
RESEARCHER(S): Wryleigh Shearin-Anderson, Mark Huff

IRB COMMITTEE ACTION: Approved
CATEGORY: Expedited

7. Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

PERIOD OF APPROVAL: February 18, 2021

Donald Sacco Jr.

**Donald Sacco, Ph.D.**
**Institutional Review Board Chairperson**

# REFERENCES


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

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.

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition*, 29, 1120–1129.

Chan, J. C., & Langley, M. M. (2011). Paradoxical effects of testing: retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 248.

LaPaglia, J. A., & Chan, J. C. (2019). Telling a good story: The effects of memory retrieval and context processing on eyewitness suggestibility. *PloS one*, *14*(2), e0212592.

Chan, J. C., & LaPaglia, J. A. (2013). Impairing existing declarative memory in humans by disrupting reconsolidation. *Proceedings of the National Academy of Sciences*, *110*(23), 9309-9313.

Chan, J. C. K., & LaPaglia, J. A. (2011). The dark side of testing memory: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied*, *17*(4), 418–432.

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

Chan J.C.K., Thomas A.K., Bulevich J.B. (2009) Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. Psychol Sci *20*(1):66–73.

Chan, J. C., Wilford, M. M., & Hughes, K. L. (2012). Retrieval can increase or decrease suggestibility depending on how memory is tested: The importance of source complexity. *Journal of Memory and Language*, *67*(1), 78-85.

Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *Quarterly Journal of Experimental Psychology*, *70*(7), 1211-1235.

Congleton, A., & Rajaram, S. (2012). The origin of the interaction between learning method and delay in the testing effect: The roles of processing and conceptual retrieval organization. *Memory & Cognition*, *40*(4), 528-539.

Dalton, A. L., & Daneman, M. (2006). Social suggestibility to central and peripheral misinformation. *Memory*, *14*(4), 486-501.

Eakin, D. K., Schreiber, T. A., Sergeant-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 813–825.

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition*, *33*(5), 770-782.

Frost, P., Ingraham, M., & Wilson, B. (2002). Why misinformation is more likely to be recognised over time: A source monitoring account. *Memory*, *10*(3), 179-185.

Frost, P., Nussbaum, G., Loconto, T., Syke, R., Warren, C., & Muise, C. (2013). An individual differences approach to the suggestibility of memory over time. *Memory, 21*(3), 408–416.

Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a self-administered interview. *Applied Cognitive Psychology*, *26*(4), 568-575.

Gabbert, F., Memon, A., & Wright, D. B. (2006). Memory conformity: Disentangling the steps toward influence during a discussion. Psychonomic Bulletin & Review, 13, 480–485.

Garcia, M., & Kornell, N. (2015). Collector [computer software].

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

Gordon, L. T., & Thomas, A. K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory & Cognition*, *42*(2), 186-197.

Gretz, M. R., & Huff, M. J. (2019). Did you wash your hands? Evaluating memory for objects touched by healthy individuals and individuals with contagious and noncontagious diseases. *Applied Cognitive Psychology*, *33*(6), 1271-1278.

Gudjonsson, G. H. (1983). Suggestibility, intelligence, memory recall and personality: an experimental study. The British Journal of Psychiatry: *The Journal of Mental Science*, 142, 35–37.

Heath, W. P., & Erickson, J. R. (1998). Memory for central and peripheral actions and props after varied post-event presentation. *Legal and Criminological Psychology*, *3*(2), 321-346.

Howe, M. J. A. (1970). Repeated presentation and recall of meaningful prose. *Journal of Educational Psychology*, 61, 214-215.

Huff, M. J., Davis, S. D., & Meade, M. L. (2013). The effects of initial testing on false recall and false recognition in the social contagion of memory paradigm. *Memory & Cognition*, *41*(6), 820-831.

Huff, M. J., & Umanath, S. (2018). Evaluating suggestibility to additive and contradictory misinformation following explicit error detection in younger and older adults. *Journal of Experimental Psychology: Applied*, *24*(2), 180.

Huff, M. J., Weinsheimer, C. C., & Bodner, G. E. (2016). Reducing the misinformation effect through initial testing: Take two tests and recall me in the morning?. *Applied Cognitive Psychology*, *30*(1), 61-69.

Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in younger and older adults: Converging evidence from the sustained attention to response task and reading for comprehension. *Psychology and Aging*, *27*(1), 106–119.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85-97.

Lane, S. M. (2006). Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(2), 199-212.

Lane, S. M., Mather, M., Villa, D., & Morita, S. K. (2001). How events are reviewed matters: Effects of varied focus on eyewitness suggestibility. *Memory & Cognition*, *29*(7), 940-947.

LaPaglia, J. A., & Chan, J. C. (2013). Testing increases suggestibility for narrative-based misinformation but reduces suggestibility for question-based misinformation. *Behavioral Sciences & the Law*, *31*(5), 593-606.

LaPaglia, J. A., Wilford, M. M., Rivard, J. R., Chan, J. C., & Fisher, R. P. (2014). Misleading suggestions can alter later memory reports even following a cognitive interview. *Applied Cognitive Psychology*, *28*(1), 1-9.

Loftus, E. F. (1979). The malleability of human memory: Information introduced after we view an incident can transform memory. *American Scientist*, *67*(3), 312-320.

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.

Meade, M. L., & Roediger, H. L. (2002). Explorations in the social contagion of memory. *Memory & Cognition*, *30*(7), 995-1009.

Numbers, K.T., Meade, M.L. & Perga, V.A. The influences of partner accuracy and partner memory ability on social false memories. *Memory & Cognition* 42, 1225–1238 (2014).

Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing. *Memory & Cognition*, *39*(1), 155-170.

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

Paz-Alonso, P. M., & Goodman, G. S. (2008). Trauma and memory: Effects of post-event misinformation, retrieval order, and retention interval. *Memory*, *16*(1), 58-75.

Pereverseff, R. S., Bodner, G. E., & Huff, M. J. (2020). Protective effects of testing across misinformation formats in the household scene paradigm. *Quarterly Journal of Experimental Psychology*, *73*(3), 425-441.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*(6002), 335-335.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255.

Roediger, H. L., Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review*, *8*(2), 365-371.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432.

Saunders, J., & Jess, A. (2010). The effects of age on remembering and knowing misinformation. *Memory*, *18*(1), 1-11.

Schwartz, S. L., & Wright, D. B. (2012). Memory conformity for new and old items with immediate and delayed testing. *Applied Cognitive Psychology*, *26*(4), 508-515.

Skagerberg, E. M., & Wright, D. B. (2009). Susceptibility to postidentification feedback is affected by source credibility. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(4), 506-523.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117.

Takarangi, M. K., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: The development of a new stimulus set. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(5), 583-590.

Thomas, A. K., Bulevich, J. B., & Chan, J. C. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language*, *63*(2), 149-157.

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

Zaragoza, M. S., Belli, R. F., & Payment, K. E. (2007). Misinformation effects and the suggestibility of eyewitness memory. *Do justice and let the sky fall: Elizabeth Loftus and her contributions to science, law, and academic freedom*, 35-63.