Episodic future thinking in eyewitnesses: An investigation of the construction of episodic future thoughts and their impact on memory accuracy for a crime

A dissertation submitted by

McKinzey G. Torrance

in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

Tufts University

May 2025

Adviser: Ayanna Thomas

Abstract

The aim of the dissertation was to understand how episodic future thinking (EFT) and interim memory retrieval may similarly impact eyewitness memory. To investigate this question, my dissertation explored three aspects of EFT. First, it investigated how EFT may impact eyewitness memory for the original event. Second, it investigated how eyewitnesses reported their phenomenological experience associated with EFT. Lastly, it explored how the focus of EFT may modulate the effects it has on subsequent memory for an event experienced before EFT. The three experiments followed the same general format: participants watched a video of a crime (original event), and either engaged in interim memory retrieval of the event, imagined an upcoming police interview about the event, or completed a filler task. Both the interim retrieval condition and EFT condition wrote down their remembered/imagined experiences. After a 24-hour retention interval, participants took a final memory test about the original event. In Experiment 2 and 3, misinformation relating to the original event was introduced to participants after the 24-hour retention interval. I found that the number of past (i.e., original event) related details generated during EFT (Experiment 2 & 3) and during interim retrieval (Experiment 1 & 2) positively predicted memory accuracy for the original event. Additionally, participants' phenomenological experience of imagining the police interview suggested that they were able to generate detail rich EFT. Lastly, the number of future (i.e., police interview) focused details produced during EFT positively predicted misinformation production (Experiment 3). Together, these experiments suggest that engaging in EFT can impact participants learning and later memory of previously presented information (i.e., original event). Further, participants may be able to richly imagine novel events for which they have limited prior direct experience. Overall,

this research suggests that people can have their memory for previously experienced events impacted by subsequent imaginings associated with that event.

Acknowledgements

Thank you to everyone who has provided me with support and guidance during these last five years. Without you, I would not be able to successfully submit this dissertation.

First, thank you to my advisor, Ayanna Thomas. I stand here at the metaphorical finish line of this milestone due to your tutelage. Thank you for the hours you dedicated to brainstorming research ideas and editing papers with me. I would not be the researcher I am today without your years of guidance. I also thank you for your friendship. I would not have eaten at nearly as many delicious restaurants without your recommendations.

Thank you to Dr. Elizabeth Race and Dr. Sara Johnson. Dr. Race, without your initial guidance in the field, and your endless feedback since, I would never have envisioned myself as a memory researcher. Dr. Johnson, through your classes and mentorship, I came to love statistics and teaching. I have grown immensely as a researcher and educator due to both of you. Thank you.

Thank you to my family. Your endless love and support fueled me through this process. A special thank you to my husband, Eric, and our cat, Opal, for your daily companionship and belief in my capabilities.

Finally, thank you to my friends who cheered me on from the beginning. To everyone in Austin, Boston, New York, and abroad, I have always felt your love. Thank you to my friends in my academic cohort and fellow PhD students. Getting to grow alongside you all and witness your passion in your own work inspired me to continue with my own.

Table of Contents

List of Tables

List of Figures

Introduction

Witness memory accuracy and reliability have garnered much attention as the consequences of error are dire. Importantly, decades of research have found that, when eyewitnesses are exposed to misinformation related to the crime, they have significantly decreased memory accuracy for the crime when later tested (for review see Loftus, 2005). Previous research has focused on the impact of externally presented misinformation (i.e., misleading police questions, inaccurate news reports, etc.), however eyewitnesses may create their own sources of alternative information by imagining original event-related future events, such as giving testimony. Imagining these types of future events is called *episodic future thinking* (EFT).

Witnesses sometimes wait days, weeks, and even months between interviews and before a court appearance. During this period, they may not only reflect upon what happened during the original event but also imagine being interviewed by investigators or testifying before a judge and jury. They could imagine what questions investigators may ask, how they may feel in these moments, and how they would respond to anticipated questions. Indeed, a study that surveyed thousands of participants several times a day for 3-14 days on what they were currently thinking about found that people imagine future events more than they think about the past (Baumeister et al., 2020). Specifically, respondents reported imagining self-relevant future events 3.5 times more than thinking about their past. Thus, eyewitnesses likely imagine future police interviews or court appearances before these events occur.

EFT relies on flexibly reconstructing prior experiences and knowledge to simulate the future (for review Schacter et al., 2017). There has been a wealth of research conducted to understand the impact of interim memory retrieval on eyewitness memory (Chan et al., 2009,

2017; Gabbert et al., 2009; Gordon & Thomas, 2014, 2017; LaPaglia & Chan, 2012, 2013; Rindal et al., 2016; Thomas et al., 2017). Thus, my dissertation sought to understand how EFT and interim memory retrieval may similarly impact eyewitness memory due to their shared reconstructive nature. To investigate this question, my dissertation explored three aspects of EFT. First, it investigated how EFT may impact memory for the crime (originally experienced event). Second, it investigated how participants reported their phenomenological experience of engaging in EFT. Lastly, it explored how the focus of EFT may modulate the effects it has on memory.

EFT and Interim Memory Retrieval

EFT may impact eyewitness memory accuracy similarly to how memory retrieval impacts eyewitness memory accuracy, because EFT relies on memory retrieval (Addis et al., 2009; Irish & Piolino, 2016; Schacter et al., 2017). Research suggests that retrieval plays an important role in long term retention of information (Arnold & McDermott, 2013; Karpicke et al., 2014a; Karpicke, 2017; Karpicke & Roediger, 2008; Rowland, 2014) and increased susceptibility to interference (Chan et al., 2009; Gordon et al., 2015; Gordon & Thomas, 2014, 2017; LaPaglia & Chan, 2013). Thus, engaging in EFT has the potential to benefit or hinder eyewitness memory for the original event.

Over a century of research has examined the benefit of memory retrieval on long term memory (for review see Karpicke, 2017). Specifically, memory retrieval can result in better long-term retention of the retrieved information compared to when information is not retrieved. This finding has been termed the Testing Effect, and one proposed mechanism for this finding is called the Episodic Context Account (Karpicke et al., 2014b). This account posits that people reinstate the episodic context associated with an item during memory retrieval. The context

representation associated with the item is then updated to include both the original context the information was learned in and information from the present retrieval context. Then, when people attempt to later retrieve the item again, the updated context of the item facilitates the retrieval of those items resulting in improved performance. Another way to conceptualize this process is as a backwards effect of testing given that the act of memory retrieval is improving the memory of previously learned information. In the context of eyewitnesses, memory retrieval prior to encountering misinformation related to the crime significantly improved memory accuracy (Gabbert et al., 2009; LaPaglia & Chan, 2012; Pansky & Tenenboim, 2011). Thus, the backward effect of testing is assumed to protect eyewitnesses from the possibly negative impact of subsequently presented misinformation.

Memory retrieval can not only result in a backwards effect of testing, but also a forward effect of testing (e.g., Test Potentiated Learning; Wissman et al., 2011). Indeed, memory retrieval prior to presentation of new information has been found to result in increased memory accuracy for the subsequently presented information. In the context of eyewitnesses, this forward effect of testing can lead to eyewitnesses becoming more susceptible to subsequently presented misinformation (e.g., Gordon et al., 2015). Indeed, when eyewitnesses engage in memory retrieval prior to encountering misinformation related to the original event, their memory accuracy for the original event can be significantly worse than participants that did not engage in memory retrieval (Chan et al., 2009, 2017; Thomas et al., 2010, 2017; Wilford et al., 2014).

The memory retrieval that is necessary to engage in EFT may lead to backward testing effects, forward testing effects, or both. I conducted two levels of analysis to investigate how EFT was impacting the learning of previously and subsequently presented information. First, I compared the impact of EFT on eyewitness memory accuracy (Experiment 1 and 2) and

misinformation production (Experiment 2) compared to conditions that engaged in interim free recall of the original event or a filler task. Second, I analyzed how the amount of retrieved content from the original event predicted memory accuracy and misinformation production. Higher memory accuracy for the original event for the EFT condition compared to the filler condition would be indicative of a backwards effect of testing. This finding would be further supported if the number of past focused (i.e., original event) details generated during EFT positively predicted final memory accuracy for the original event. A forward effect of testing would be seen if EFT is associated with higher misinformation production on the final memory test compared to the filler task. This result would be further supported if the number of past focused details generated during EFT positively predicted misinformation production. These results would suggest that retrieved original event details resulted in test potentiated learning of those details (Gordon et al., 2015; Gordon & Thomas, 2014, 2017).

The construction of EFT for novel events

Many studies have suggested that overlapping brain regions are associated with episodic remembering and EFT (for review see Schacter et al., 2017; Szpunar, 2010). A major takeaway from this body of research is that when one struggles or fails to remember the past, they also struggle or fail to imagine the future. Researchers suggested that EFT relies on flexibly retrieving and combining elements of past experiences into novel events that could occur (i.e., constructive episodic simulation hypothesis, Schacter & Addis, 2007). Importantly, other researchers have found that people not only rely on their episodic memories but also their semantic memories to scaffold their episodic future thoughts (Irish et al., 2012). Semantic memory represents the general knowledge one holds about the world. For example, if one was to imagine a kitchen, they would know whether to include a refrigerator or a bed in the room given their semantic memory

of what objects should be in a kitchen. When exploring the role of semantic memory in EFT, patients with Semantic Dementia were found to have similar episodic thinking deficits to patients with Dementia of the Alzheimer's Type (DAT), even though they had intact episodic memories (Irish et al., 2012). These findings support the semantic scaffolding hypothesis, which posits that semantic memory provides an essential framework for retrieval of past events and EFT (for review see Irish & Piolino, 2016).

Eyewitnesses to a crime present an interesting context to explore EFT given that witnessing a crime and all subsequent related events are likely novel. That said, the average American may have a wealth of semantic information of crimes and the criminal justice system garnered from media (e.g., film, serialized episodes, podcasts, and news articles covering real-world trials). While they may also have episodic memories relating to certain media (e.g., memory of a specific scene in a movie), these sources of information could also create semantic memories of how these situations are expected to function. Research with DAT patients demonstrates that when people have limited access to episodic memories, and/or have disrupted episodic memory processing, they struggle to form EFT (Addis et al., 2008, 2009). Thus, participants might struggle to generate EFT if they do not have related episodic memories to rely on.

My dissertation explored how neurotypical participants may be able to successfully engage in EFT of activities that follow witnessing a crime. While episodic memories may be less available for scaffolding EFT, semantic memories are likely available to simulate a possible future. The construction of EFT was assessed in the written descriptions participants gave when engaging in EFT as well as the phenomenological ratings associated with those simulations. Research has consistently demonstrated that experienced events are associated with more

sensory characteristics than imagined events (D'Argembeau & Van der Linden, 2004, 2006; Johnson, 2006; Johnson et al., 1988). Additionally, future events have been shown to have a positivity bias (Newby-Clark & Ross, 2003) and are associated with cognitive operations, such as reasoning, searching memory, and imagery processes (Johnson et al., 1981; McDonough & Gallo, 2010). Together, these findings suggest that past and future events have significant differences in how they are experienced and described, suggesting a means to differentiate between events via their phenomenological characteristics. Therefore, I assessed phenomenological characteristics of EFT and retrieved memories in all experiments (for review see Johnson, 2006).

The Construction of EFT and Memory Accuracy

Phenomenological experience, or the memory characteristics associated with the event, not only demonstrate how people differentially experience events, but also have been found to aid memory retrieval. Indeed, past research has found that when event characteristics become more similar to each other, people experience more reality monitoring errors. That is, they are more likely to attribute an imagined event or action to a perceptual experience (Johnson, 2006). Importantly, while experienced events and imagined events are often associated with different memory characteristics, there are important circumstances when the memory characteristics associated with an imagined event can become more similar to an experienced event. Past research has found that temporal distance to the present (D'Argembeau & Van der Linden, 2004), familiarity with the location (Szpunar & McDermott, 2008), and valence of the event (Painter & Kring, 2015; Rubin, 2014) are all factors that can lead to simulations with memory characteristics that are more similar to those of experienced events. Thus, participants may have less vivid or sensory rich memory characteristics associated with their imagined police

interviews compared to when they remember the original event. Specifically, eyewitnesses may have little to no prior experiences being interviewed by police to aid in the simulation of the interview (Schacter et al., 2017). That said, they may be able to use their semantic memory of police interviews to aid in simulating the event (Irish & Piolino, 2016). This may lead to similarly strong memory characteristics associated with imagined and remembered events. The construction of EFT will be assessed by analyzing the free-responses participants generate when engaging in memory retrieval or EFT, as well as by analyzing memory characteristic ratings participants give related to their memory retrieval or simulation experience.

My dissertation explored the construction of EFT and its impact on memory accuracy when the focus of EFT is either explicitly manipulated or not. In Experiments 1 and 2 participants engaged in EFT with little constraints. Specifically, they were instructed to imagine the upcoming police interview and how they would be questioned about the crime. In Experiment 3, I manipulated what participants were asked to focus on during EFT (i.e., perceptual or aperceptual details). Specifically, participants were directed to focus on the perceptual experience of the imagined police interview (e.g., what they would see, hear, smell, touch) or the aperceptual experience of the imagined police interview (e.g., what emotions they would feel). Past research has found that when participants were asked to focus on aperceptual details (i.e., thoughts and emotions) while imagining a future event or remembering a past event, there was a decrease in the salience of context and sensory characteristics over time compared to participants that focused on perceptual details (i.e., sensory details, Suengas & Johnson, 1988). Experienced and imagined events became more similar in the subjective number of thoughts and feelings over time. Experiment 3 explored how focusing on perceptual and aperceptual details

during EFT impacted how EFT was experienced, and the downstream consequences of that focus on subsequent original event memory.

My dissertation also explored how EFT impacts eyewitness memory. Because successful EFT requires retrieval from past experiences, I hypothesized that EFT would influence subsequent memory for an original event in ways similar to interim explicit retrieval. Consistent with prior interim retrieval research, I expected engaging in EFT to result in forward and backward effects of testing. I further investigated the phenomenological experience of EFT for an unlikely experienced event (i.e., police interview), and how this experience compared to the phenomenological experience of remembering an experience event (i.e., original event) and imagining a previously experienced event (i.e., birthday). Together, these analyses provided insight into how EFT is experienced for participants and how this experience could influence the processing for previously encountered (i.e., original event) and subsequently encountered (i.e., misleading police interview questions) information. Lastly, I directly manipulated the focus of EFT (i.e., focus on perceptual or aperceptual details) to examine how the content of EFT may further influence how EFT impacts eyewitness memory.

Experiment 1

Experiment 1 contained two sessions. In the first session all participants watched a crime video (original event) and either engaged in interim retrieval for the original event, imagined a police interview (EFT), or completed a filler task. In session two, all participants engaged in a final free recall test of the original event and answered questions relating to their prior experiences with eyewitness related media.

The first aim of Experiment 1 was to compare the underlying processes of memory retrieval and EFT. Specifically, I investigated how engaging in EFT improved memory for the

original event as compared to interim memory retrieval. I further examined how the contents of EFT and interim memory retrieval of the original event predicted later memory accuracy for the original event. Through these analyses I directly explored how the retrieval process used during EFT impacted eyewitness memory and if the impacts of EFT on eyewitness memory for the crime were comparable to the impacts of interim memory retrieval. Prior interim memory retrieval research has demonstrated that interim memory retrieval can improve memory accuracy for retrieved information due to a backward effect of testing (Karpicke, 2017; Karpicke & Roediger, 2008; Rowland, 2014). Importantly, memory retrieval of prior experiences is critical to engage in EFT (Addis et al., 2008; Irish et al., 2012; Schacter et al., 2017). Thus, I hypothesized that participants who engaged in interim memory retrieval or EFT would have higher memory accuracy for the original event compared to those who engaged in a filler task. I further expected that the number of original event specific details that one generated during interim memory retrieval and EFT would positively predict memory accuracy for the original event.

The second aim of the study was to examine how the phenomenological experience of generating an event varied across interim task (interim retrieval, EFT, filler) and between imagined/remembered events. Specifically, participants answered Memory Characteristics Questionnaire (MCQ) questions after each remembered/imagined event. The MCQ is a questionnaire that measures the phenomenological experiences one experiences when imagining or remembering an event (Johnson et al., 1988). Prior research found that experienced events were often rated with higher sensory characteristics compared to imagined events (for review see Johnson, 2006) and that prior experiences can lead to higher ratings in phenomenological characteristics (e.g., D'Argembeau & Van der Linden, 2004). Thus, Experiment 1 aimed to understand how participants engaged in EFT for a police interview, and if this experience

differed from the experience of remembering the original event or imagining a previously experienced event. Given that participants likely have little to no prior experience being interviewed by police, I hypothesized that they would rate the experience of imagining a police interview as having less clear sensory characteristics compared to remembering the original event and imagining a previously experienced event. This aim provided insight into how people engage in EFT when they possibly have few episodic memories related to being an eyewitness. I further analyzed if prior experience with eyewitness related media impacted EFT to examine if prior experiences could predict one's experience imagining an event.

Method

Participants

In the first session of Experiment 1, 170 participants were sampled from Prolific, an online platform for recruiting participants. There were six participants that failed the attention checks which were designed to assess participant's comprehension of task instructions (see Appendix A). Thus, 164 participants were invited to participate in session 2. Session 2 was 24 hours after session 1. There were 135 participants that participated in session 2 (~20% attrition rate). During the scoring of session 2 data, five participants were found to write their final memory test about a crime that they did not experience during the experiment, and they were removed from analysis. This resulted in a final sample of 131 participants for analysis, with 45 participants in the EFT condition, 39 in the initial retrieval condition, and 47 participants in the filler condition.

Participants were between the ages of 18-35 (M = 26.93, SD = 4.92). Participants identified as women (n = 71), men (n = 41), non-binary (n = 11), or as trans-men (n = 2). Participants identified as White (n = 65), Black (n = 25), having multiple racial identities (n =

13), East Asian (*n* = 10), Latino / Hispanic (*n* = 7), Southeast Asian (*n* = 6), Native American / Alaskan Native (*n* = 2), Middle Eastern (*n* = 2), or South Asian (*n* = 1). Participants reported varied levels of education. Participants reported attending some high school (*n* = 3), receiving a high school diploma (*n* = 17), attending some college (*n* = 31), receiving an Associate's degree (*n* = 12), receiving a Bachelor's degree (*n* = 55), or receiving a Master's degree or above (*n* = 13). Participants reported having various levels of English exposure before the age of 5 with the majority reporting they were only exposed to English at home (*n* = 91). Participants also reported being exposed to English and other languages at home through age 5 (*n* = 39) or had no regular English exposure through age 5 (*n* = 1).

Forty-six participants reported witnessing a crime before. I calculated all of the analyses with and without these participants and found that the pattern of results was consistent.

To assess if the analyses were appropriately powered, four separate sensitivity analyses (Giner-Sorolla et al., 2024) were conducted. This allowed me to assess the smallest detectable effect size given our sample size. All sensitivity analyses used a power of .8. Using G*Power, I calculated the smallest observable effect size for the effect of condition (interim retrieval, EFT, filler) on memory accuracy for the original event (R^2 = 0.063). I calculated the effect size with a sample size of 117, given that G*Power assumes equal size groups and the lowest sample size was 39 (interim retrieval). The effect sizes observed within the experiments were larger than the smallest detectable effect size. I then calculated the smallest effect size for the effect of retrieval focused details on memory accuracy in the retrieval condition (R^2 = 0.175). The effect size was larger than the smallest detectable effect size. Next, I calculated the smallest effect size for the effect of retrieval focus and future focused details on memory accuracy in the EFT condition (R^2 = 0.187). The effect size was smaller than the smallest detectable effect size. Lastly, I calculated

the smallest effect size for the effect of condition (interim retrieval, EFT, filler) on the ratings of the MCQ questions (η^2 = 0.005). I calculated the effect size with a sample size of 117. The effect sizes of each one-way ANOVA analyzing the impact of condition on MCQ question were smaller than the smallest detectable effect size, with the exception of the analyses of the MCQ question relating to the event being experienced like a story.

Materials

Original Event. The original event depicted a non-violent theft where a man steals a purse and the victim chases after him (Greene et al., 2022). It was approximately 2 minutes in length. There was no audio that accompanied the video.

Common Event Trial. All participants completed a common event trial and were asked to remember their last birthday or imagine their next birthday. Participants in the interim retrieval condition had a retrieval practice trial, participants in the EFT condition had an imagination trial, and participants in the filler condition were randomly assigned to one of the two practice trial types. Participants had three minutes to engage in either type of trial (See Appendix B). This trial allowed participants in the interim retrieval and EFT conditions to become accustomed to the imagination/memory retrieval instructions prior to either remembering the original event or imagining the police interview. Furthermore, it created a comparison event for the EFT condition. Specifically, I was able to compare the phenomenological experience of imagining a prior experienced event (birthday) to an event one had not likely experienced before (police interview).

EFT. Participants in the EFT condition were told that they were contacted by the police to give a statement about what they saw during the crime tomorrow. They were asked to imagine what giving their statement to the police would be like in as much detail as possible (see

Appendix C). They had six minutes to imagine giving their statement to the police. They were not able to leave the webpage before the six minutes were up and were encouraged to add more information to what they imagined would happen if they finished writing before the six minutes were up. There was a timer at the top of the webpage counting down the six minutes so participants would be aware of how much more time they had to add more information.

Interim Retrieval. Participants in the interim retrieval condition were told that they were contacted by police to give a statement about what they witnessed. They were prompted to remember the crime in as much detail as possible (see Appendix D). This was a free recall task for the original event. They had six minutes to remember as much about the original event as possible. They were not able to leave the webpage before the six minutes were up and were encouraged to add more information to what they remembered if they finished writing before the six minutes were up. There was a timer at the top of the webpage counting down the six minutes.

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

Final memory test. The final memory test was a free recall memory test for the original event (see Appendix E). Participants had six minutes to remember as much detail about the

original event as possible. They were not able to leave the webpage before the six minutes were up and they were encouraged to add more information to what they remembered if they finished writing before the six minutes were up. There was a timer at the top of the webpage counting down the six minutes. While the presence of a timer may have encouraged a sense of urgency, I do not believe this would confound our results given that reporting eyewitness testimony likely has a sense of urgency in the real world. Furthermore, this time was present across all conditions resulting in a consistent sense of urgency which would prevent confounds across experimental conditions.

Memory Characteristics Questionnaire. The Memory Characteristics Questionnaire (MCQ) was originally created by Johnson and colleagues (1988). The present study used a variation of this questionnaire that was used by Rubin in a study of valence and EFT (2014). The present study used questions from the MCQ that had participants rate their experience of engaging in EFT (e.g., clarity, valence, and modality of events). All questions were measured on a Likert scale (See Appendix F).

Sudoku. Another filler task was two games of Sudoku for participants to complete at their own pace. There was a timer counting down 10 minutes at the top of the web page. Participants were automatically moved to the next section of the experiment after the 10 minutes were up. Even if participants finished the Sudoku puzzles early, they were not allowed to move to the next section until the timer reached zero.

Attention check question. The attention check question asked participants if they had died of a heart attack while watching the original event. Participants could respond yes, no, or maybe, and only participants that reported no were counted as passing the attention check.

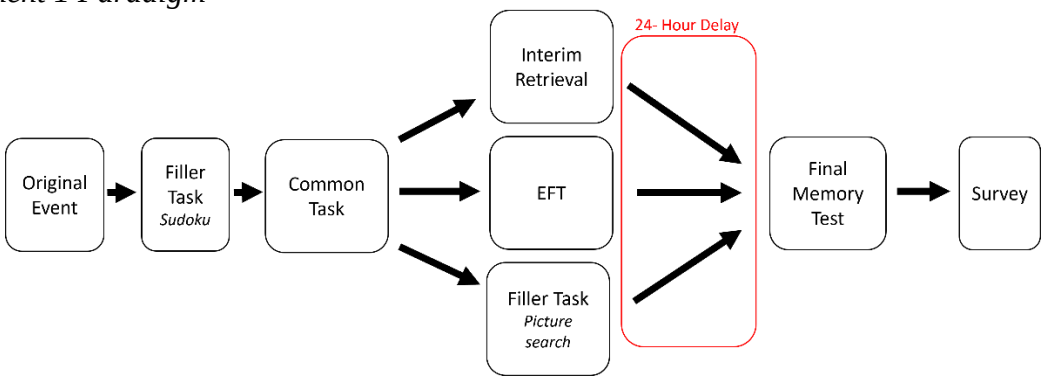
Prior eyewitness experience survey. All participants completed the prior eyewitness

media survey. This survey asked if participants had witnessed a crime before (yes, no). It also asked participants to tell approximately how many original event related videos (TV shows, movies, YouTube videos), podcasts (e.g., true original event podcasts), and news articles they had encountered in their lives. This survey was designed to assess if the number of prior experiences with eyewitness related media impacted how eyewitnesses engaged in EFT when imagining a police interview.

Procedure

Figure 1

Experiment 1 Paradigm



Participants began the experiment by giving their informed consent. They then watched the original event (see Figure 1). After the original event, they answered an attention check question. Then, they completed a filler task which was sudoku for 10 minutes. Following the filler task, participants completed a commen event task of remembering a past event or imagining a future event. Participants in the filler condition completed the common event task to ensure they took about the same amount of time to complete the experiment. After, participants either completed a second filler task (i.e., picture search), engaged in EFT, or engaged in interim retrieval of the original event. If participants engaged in interim retrieval or EFT, they rated their experience of EFT / interim retrieval via the MCQ. All participants returned after a 24-hour delay. The 24-hour delay allowed time for the imagined event to be consolidated in memory

(McGaugh, 2000). Upon return, participants took a final memory test relating to the original event and answered MCQ questions again. Following the final memory test, participants answered demographic questions, as well as a survey relating to their prior experience with eyewitnesses and eyewitness related media.

Scoring

Participants interim tasks (remembering the original event or imagining the police interview) were scored by two research assistants. Each research assistant was unaware of the hypotheses of the experiment. The interim tasks were scored for the number of reported details that focused on past information (from the original event) and future information (related to the police interview). The scoring procedure was based on the cognitive interview scoring technique developed by Memon and colleagues (1997) where sentences were broken up into four different detail types; people, actions, objects, and location. Thus, for the phrase "I saw a woman walking down the street with her baby in the stroller" there are three people details (I, woman, baby), two action details (saw, walking), one location detail (the street), and one object (stroller). Adjectives were also counted as details. For example, "I saw a woman with *brown hair* walking down the street with her baby in the stroller" there would be two extra scored details, one being hair and another being for the adjective describing the hair. These details were coded so that scoring could capture when a participant included more specificity in their descriptions. Details were determined to be past or future oriented based on the tense of the verb that was used, and if participants were referencing the original event or imagining the police interview. If participants started to reference events that were not present within the study, these statements were not coded as future or past details. These statements were not coded because they did not relate to the event participants experienced in the past (the original event) or the event they would

experience in the future (the police interview). Furthermore, participants interjections were not scored as past or future details (i.e., “It was stupid of the mother to leave her child”).

Scoring resulted in two values: the number of future focused details generated during the event and the number of past focused details generated during the event. I evaluated interrater reliability through having a second rater score a subset (half) of all participant responses. Interrater reliability for the scoring of past oriented details during interim retrieval of the original event was high (r = 0.96). No participants reported future oriented details in the interim retrieval condition. The interrater reliability for future details (r = 0.91) and past oriented details (r = 0.98) in the EFT condition were also high. For the final score for each participant, the score from the primary rater of each section was retained instead of averaging scores of two raters. I took this approach given that both raters did not score every participant which would lead to only a subset of participants having an average score.

The final police report of the original event was scored for correct and incorrect details using the same methods as above. Specifically, people, action, location, object, and adjectives were coded as either correct (as occurring within the original event) or incorrect (not occurring within the original event). Research assistants both scored 10 responses from each condition for accurate and inaccurate details. There was high interrater reliability for accurate details (r = 0.96) and inaccurate details (r = 0.98).

Results

All data were deemed suitable for the planned analysis (See Appendix G for complete data screening).

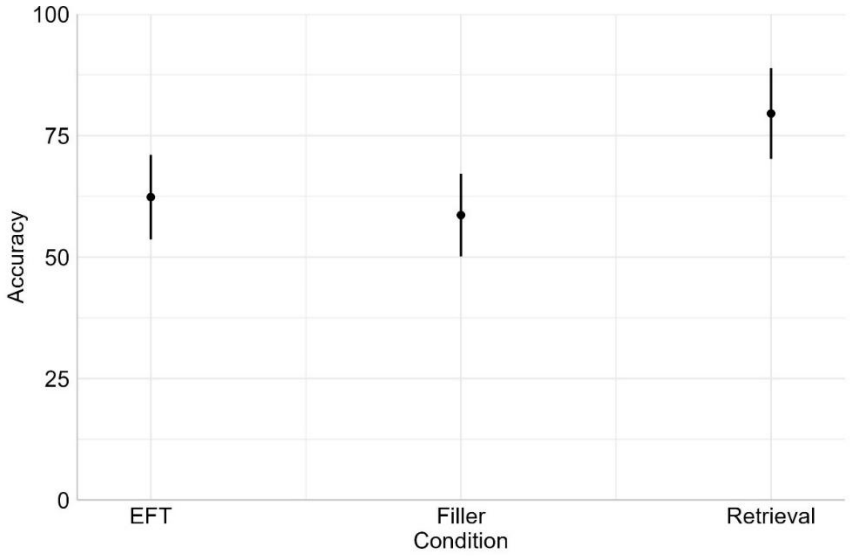
The impact of the interim task on memory of the original event

I estimated a linear regression to examine the impact of interim task (interim retrieval, EFT, filler) on memory accuracy for the original event on the final memory test. Memory accuracy was a continuous outcome variable measuring the number of correct details produced during the police interview. Interim task was a categorical variable with three groups. The referent group in the model was the filler group.

Participants that engaged in interim retrieval produced significantly more correct details about the original event in the final memory test compared to participants that completed a filler task. ($\beta = 20.91$, $p = .001$, $R^2 = 0.08$; see Figure 2). There was not a significant difference in the number of correct details reported in the final memory test between the filler and the EFT conditions ($\beta = 3.70$, $p = .549$).

Figure 2

Experiment 1 Effect of Interim Task on Final Memory Test Accuracy



Note. Dots represent model predicted means and lines represent the 95% confidence interval around the mean for each condition. Accuracy is measured as the number of correct details produced on the final memory test.

The role of event construction in memory accuracy for the original event

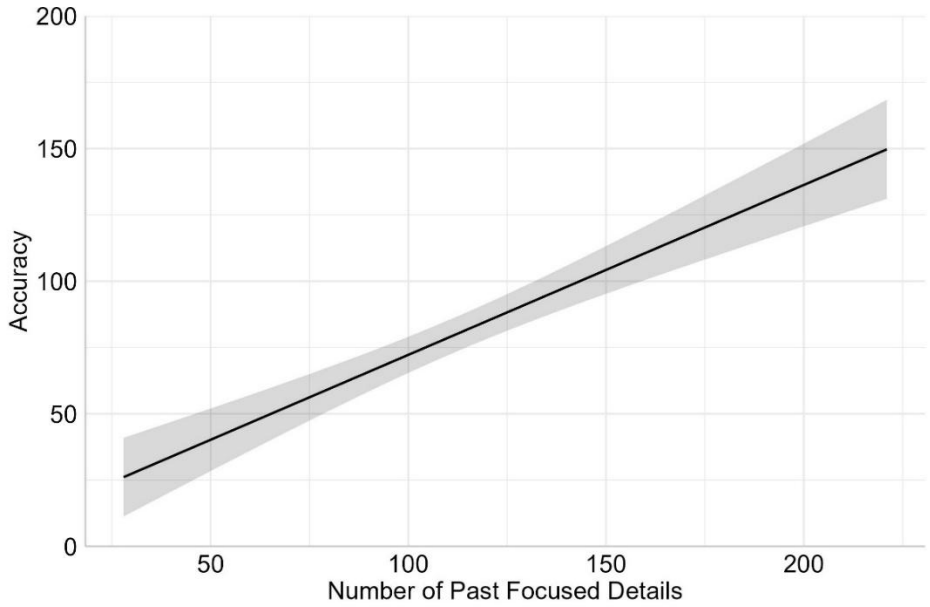
I estimated a linear regression exploring how the construction of EFT and interim retrieval impacted later memory for the original event. Importantly, when analyzing the results from past and future event constructions, participants in the interim retrieval condition did not report any future oriented details. Thus, the role of event construction on memory accuracy for the original event was investigated in two separate linear models; one for the interim retrieval condition and one for the EFT condition.

The impact of event construction on memory accuracy for the original event in the interim retrieval condition

I estimated a linear regression with a single continuous predictor variable of past focused details on the outcome variable of number of correct details produced about the original event during the final memory test. First, I looked at only participants in the interim retrieval condition. The number of past focused details significantly predicted later memory accuracy for the original event ($\beta = 0.64$, $p < .001$, $R^2 = 0.64$; see Figure 3).

Figure 3

Experiment 1 Association Between Final Memory Test Accuracy and Number of Past Focused Details Generated during Interim Retrieval



Note. The line represents the model predicted relationship while the area around line represents the 95% confidence interval. Accuracy is measured as the number of correct details produced on the final memory test.

The impact of event construction on memory accuracy for the original event in the EFT condition

I estimated a linear regression with two continuous predictor variables (number of past focus details and number of future focused details) on the outcome variables of number of correct details produced on the final memory test. The participants included in this analysis were those in the EFT condition. The number of past focused details (β = 0.27, p = .079) and the number of future focused details (β = 0.14, p = .156) did not significantly predict memory accuracy for the original event (R^2 = 0.09).

Phenomenological experience (MCQ) constructing events

I conducted three separate analyses of MCQ responses. First, was the comparison of MCQ responses for the experience of remembering the original event during the police interview across the interim task conditions (interim retrieval, EFT, filler). I was able to assess if there were any differences in the perceived experience of remembering the original event due to

interim task. The second analysis was of MCQ responses for when participants engaged in EFT or interim retrieval of the original event. With this analysis, I assessed how participants experience remembering a recently experienced event may be different than generating a novel event that is unrelated to prior experiences. The final analysis was a comparison of MCQ responses when participants engaged in EFT for a police interview and EFT for their future birthday. I assessed if participants phenomenological experience of imagining a police interview was different than imagining an event which one has many prior experiences with.

Impact of interim task on the phenomenological experience (MCQ) of remembering the original event

I conducted nine one-way ANOVAs analyzing the impact of interim task (interim retrieval, EFT, filler) on the phenomenological experience of the final memory test. There were no statistically significant differences in the phenomenological experience of remembering the original event during the final memory test between the three conditions (see Table 1).

Table 1

Experiment 1 Comparisons of Final Memory Test MCQ Ratings Between Interim Tasks

MQC Question	F-Value	p-value	η^2
Reliving	1.97	.143	0.03
Intensity	0.49	.613	0.01
Reaction	0.49	.615	0.01
Valence	2.07	.130	0.03
See	2.61	.078	0.04
Setting	2.81	.064	0.04
Hear	2.92	.057	0.04
In Words	1.28	.282	0.02

Story †	0.28	.758	0.004

Note. † Denotes the ANOVA that did has an effect size larger than the minimal detectable effect size determined by a sensitivity analysis.

Differences in the phenomenological experience (MCQ) of interim retrieval and EFT of the police interview

I conducted nine t-tests evaluating the phenomenological differences in the experience of engaging in interim retrieval versus EFT of the police interview (See Table 2). Participants in the interim retrieval condition rated the setting of the event clearer (*M* = 5.74*)* compared to participants in the EFT condition (*M* = 5.18). Participants in the interim retrieval condition rated the visual clarity of the event as clearer (*M* = 5.97) compared to participants in the EFT condition (*M* = 5.11). Participants in the interim retrieval condition rated the event as more story-like (*M* = 5.74) compared to participants in the EFT condition (*M* = 5.11). Lastly, participants in the EFT condition reported the event had stronger audio clarity (*M* = 4.29) compared to participants in the interim retrieval condition (*M* = 2.87).

Table 2

Experiment 1 Comparisons of MCQ Ratings of Interim Retrieval and EFT of the Final Memory Test

MCQ Question	$M_{interim\ retrieval}$	M_{EFT}	t-value	p-value	*d*
Reliving	5.28	5.00	0.91	.365	0.19
Intensity	3.51	3.42	0.25	.802	0.06
Reaction	2.33	2.62	-0.81	.418	-0.18
Valence	-0.62	-0.49	-0.61	.545	-0.13
See	5.97	5.11	3.68	<.001*	0.79
Setting	5.74	5.18	2.03	.045*	0.45
Hear	2.87	4.29	-3.41	.001*	-0.76

In Words	4.44	4.42	0.04	.972	0.01
Story	5.74	5.11	2.39	.019*	0.51

Note. * Denotes p < .05

Differences in the phenomenological experience (MCQ) of EFT

I conducted nine t-tests assessing if the phenomenological experience of imagining a future birthday was different from imagining a future police interview (See Table 3). The only statistically significant phenomenological difference in the experience of imagining these two events was in their valence. Specifically, participants rated their imagined birthday with a more positive valence (M = 1.22) compared to their imagined police interview (M = -0.49).

Table 3

Experiment 1 Comparisons of MCQ Ratings of Imagining a Future Birthday and Imagining the Police Interview

MCQ Question	$M_{birthday}$	$M_{police\ interview}$	t-value	p-value	d
Reliving	4.58	5.00	-1.33	.186	-0.28
Intensity	3.36	3.42	-0.19	.849	-0.04
Reaction	2.47	2.62	-0.45	.656	-0.09
Valence	1.22	-0.49	6.37	<.001*	1.34
See	5.04	5.11	-0.23	.819	-0.05
Setting	5.38	5.18	0.77	.446	0.16
Hear	3.98	4.29	-0.86	.391	-0.18
In Words	4.20	4.22	-0.63	.534	-0.13
Story	4.89	5.11	-0.71	.481	-0.15

Reported prior experience with eyewitnesses

To understand if participants' prior exposure to police related media impacted their memory accuracy for the original event or their experience of imagining the police interview, I estimated several linear regressions. The three continuous predictor variables were the number of prior video-based eyewitness media they had watched (i.e., YouTube videos, movies, TV shows), the number of prior news media they had encountered (i.e., news articles related to court cases), and the number of podcasts or auditory media they had heard (i.e., true crime podcasts, radio news). The outcome variables for the separate models were correct details given about the original event during the final memory test, MCQ ratings of visual clarity during EFT for the police interview, the setting of the police interview, audio clarity during the police interview and how story-like it was to construct the police interview. Prior exposure to police-related media did not predict memory accuracy of the original event, nor did it predict the phenomenological experience of imagining the police interview (See Tables 4-8).

Table 4

Experiment 1 Linear Model Results for Relationship between Prior Media Exposure and Final Memory Test Accuracy

Prior Experience Questions	β (SE)	*p*-value
Prior Video Exposure	0.04 (0.38)	.912
Prior Article Exposure	-0.01 (0.26)	.967
Prior Audio Exposure	0.02 (0.03)	.597

Table 5

Experiment 1 Linear Model Results for Relationship between Prior Media Exposure and Visual Experience (MCQ) of EFT Police Interview

Prior Experience Questions	β (SE)	*p*-value

Prior Video Exposure	-0.002 (0.01)	.904
Prior Article Exposure	0.007 (0.01)	.553
Prior Audio Exposure	-0.00007 (0.001)	.958

Table 6

Experiment 1 Linear Model Results for Relationship between Prior Media Exposure and Clarity of Setting (MCQ) of EFT Police Interview

Prior Experience Questions	β (SE)	*p*-value
Prior Video Exposure	-0.004 (0.02)	.824
Prior Article Exposure	-0.001 (0.01)	.957
Prior Audio Exposure	0.001 (0.001)	.412

Table 7

Experiment 1 Linear Model Results for Relationship between Prior Media Exposure and Clarity of the Audio Experience (MCQ) of EFT Police Interview

Prior Experience Questions	β (SE)	*p*-value
Prior Video Exposure	-0.003 (0.02)	.896
Prior Article Exposure	0.011 (0.02)	.522
Prior Audio Exposure	0.001 (0.002)	.768

Table 8

Experiment 1 Linear Model Results for Relationship between Prior Media Exposure and the Story-Like Experience (MCQ) of EFT Police Interview

Prior Experience Questions	β (SE)	*p*-value
Prior Video Exposure	-0.009 (0.02)	.646
Prior Article Exposure	0.01 (0.01)	.459
Prior Audio Exposure	0.0001 (0.002)	.996

Experiment 1 Discussion

In Experiment 1, I investigated if the memory retrieval used to construct EFT impacted memory for the original event. Specifically, I compared the impact of EFT on memory accuracy to the impact of interim retrieval on memory accuracy. Interim retrieval of original event details resulted in better memory accuracy for those details compared to completing the filler task. Further, the number of past details generated during interim retrieval positively predicted participant memory accuracy for the original event during the final memory test. These findings are consistent with the Testing Effect literature that suggests that engaging in an interim memory test can positively impact long term memory for retrieved items (Karpicke, 2017; Karpicke et al., 2014a; Karpicke & Roediger, 2008; Rowland, 2014).

Importantly, engaging in EFT did not impact participant memory accuracy for the original event compared to completing the filler task. Further, the number of past details and future details generated during EFT were not statistically significant predictors of final memory accuracy. It was hypothesized that the memory retrieval utilized during EFT would lead to similar backward testing effects on eyewitness memory as seen in the interim retrieval condition. This would be observable as higher memory accuracy for the original event compared to the filler condition and a positive relationship between the number of past focused details generated during EFT and final memory accuracy for the original event. Thus, these results suggest that the EFT does not result in a Testing Effect. Of note, the analysis of the relationship between the number of past focused details produced during EFT on memory accuracy for the original event was underpowered. Thus, Experiment 1 may not be appropriately powered to detect the effect of past focused or future focused details during EFT on memory accuracy for the original event.

The second goal of Experiment 1 was to examine how participants engaged in EFT for a police interview. To conduct this analysis, I examined phenomenological experiences of both imagined and remembered events via MCQ ratings. In the first analysis, interim tasks were not found to impact participant's phenomenological experience (MCQ) of remembering the original event during the final memory test. These results suggest that interim tasks may result in variation in memory accuracy, but not in differences in phenomenological experience of remembering the original event.

In the second analysis I compared the phenomenological experience of engaging in interim retrieval of the original event and EFT of the police interview. Interim retrieval of the original event was rated with stronger ratings of the clarity of setting, overall visual clarity, and the event being experienced more like a story compared to EFT of the police interview. These findings are consistent with prior research that found that experienced events had higher ratings on sensory details compared to imagined events (D'Argembeau & Van der Linden, 2004, 2006; Rubin, 2014). Interestingly, participants rated the auditory experience of imagining the police interview as clearer compared to interim memory retrieval of the original event. This may be because the original event video did not include any audio. Alternatively, the police interview was specifically an imagined conversation between the participant and an investigator. Given that the imagined police interview requires participants to imagine a conversation, participants may focus on simulating audio information which could lead to higher perceived ratings in how participants "hear" the constructed event.

In the final analysis of the MCQ data, I compared the phenomenological experience of EFT for the police interview and EFT for a future birthday. Prior research suggests that people should be able to more robustly engage in EFT when they have prior experiences related to the

imagined event (Addis et al., 2008; Schacter & Addis, 2007). Given that people have not likely been questioned by police, or they have at least not experienced this event as frequently as experiencing their birthday, I expected people to imagine their birthday with more sensory details compared to the police interview. Interestingly, the only difference in the phenomenological experience of imagining these two events was in their experienced valence. People rated their imagined birthday as significantly more positive than the imagined police interview. These results suggest that participants were able to imagine a police interview as robustly as their own next birthday. One possibility is that Americans hold robust semantic knowledge about how eyewitness interviews are conducted. This semantic knowledge could be used as a scaffold to develop an EFT that is just as robust as a simulation of a future event associated with robust episodic memories. This finding would be consistent with the scaffolding account of EFT which found that semantic memories are also critical to the construction of episodic future thoughts (Irish et al., 2012; Irish & Piolino, 2016).

Another possibility is that people may use episodic memories associated with eyewitness media to engage in EFT of the police interview. Thus, participants might have episodic and semantic memories that they are using to construct their EFT. Interestingly, prior exposure to police related media did not seem to predict memory accuracy for the original event or the strength of phenomenological characteristic ratings. These results suggest that participants' prior media exposure was not directly impacting their phenomenological experience of imagining the police interview.

Experiment 2

The misinformation effect is the finding that when eyewitnesses are exposed to misleading post-event information relating to the crime, that their memory accuracy for the crime

is significantly reduced (for review see Loftus, 2005). Research suggests that repeated retrieval of the original event can impact susceptibility to misinformation and memory for the original event (for review see Chan et al., 2017). This increased susceptibility to misinformation compared to those who do not engage in interim memory retrieval is termed *retrieval enhanced suggestibility* (i.e., RES; Chan et al., 2009). One suggested mechanism behind the RES effect is that it is the result of a forward effect of testing. Forward effect of testing posits that interim memory retrieval improves the learning of the subsequently presented information (Gordon & Thomas, 2014, 2017; Pastötter & Bäuml, 2014). Importantly, interim memory retrieval does not always result in a forward effect of testing. Indeed, prior research has found that interim memory retrieval can protect against subsequent misinformation exposure compared to if someone does not engage in interim memory retrieval (Gabbert et al., 2009; LaPaglia & Chan, 2012; Pansky & Tenenboim, 2011). Memory retrieval has been found to be a key component of EFT, such that without memory retrieval one cannot engage in EFT (Addis et al., 2008; Irish et al., 2012). The aim of Experiment 2 was to examine how EFT impacts the processing of subsequently presented information.

Given that EFT in part relies on memory retrieval, engaging in EFT may similarly produce a backward or forward effect of testing. In the present study both interim retrieval and EFT are freely generated. Past research has found that free recall of an event can result in RES on a final free recall test and on a final cued recall test (Wilford et al., 2014). Thus, I predicted that interim retrieval and EFT would result in RES. Importantly, prior work suggested that interim retrieval results in RES because information retrieved during interim retrieval directs attention towards related (and possibly misleading) information encountered in the post-event information (Gordon & Thomas, 2014, 2017). I expected there to be a forward effect of testing,

if participants retrieve details from the original event during interim retrieval or EFT that are related to subsequently encountered misleading police questions. This would be demonstrated as lower memory accuracy for the original event for the interim retrieval and EFT conditions when compared to the filler condition. Conversely, if participants remember original event details during interim retrieval or EFT that do not overlap with misleading police questions, then I would not expect there to be a RES effect for these conditions compared to the filler condition. There could still be a possible backward testing effect, where participants have higher memory accuracy for the original event overall compared to a filler condition.

Experiment 2 also aimed to extend the MCQ findings from Experiment 1. Past research found that people more vividly imagine events that are based on prior experiences (e.g., D'Argembeau & Van der Linden, 2012). Thus, participants should be able to imagine their future birthday with more sensory details compared to a police interview. In Experiment 1, the phenomenological experience of imagining a birthday may not have been significantly different from imagining a police interview because participants had less time to imagine their future birthdays compared to the police interview. In Experiment 2, participants were given equal amounts of time in the EFT condition to imagine their next birthday and imagine a police interview.

Method

Participants

The sample size for Experiment 2 was decided by an *a priori* power simulation. I used simr (Green & MacLeod, 2016) to simulate the power for the effect of interim task (interim retrieval, EFT, filler) on memory accuracy for the original event at various sample sizes. I used a sample of 83 participants (interim retrieval = 27; EFT = 26; Filler = 30) to calculate an effect

size for the effect of interim task on memory accuracy. It was determined that a sample size of at least 420 was needed to power the generalized multi-level model for the effect of interim task with a power of 80.90 (CI 95% [78.32,83.29]).

Six hundred and twenty participants were sampled for the study. As in Experiment 1, participants were not invited to session 2 of the study if they failed any of the attention checks (See Appendix A). Five hundred and forty-seven participants were invited to take part in session 2 of the experiment. There was an attrition rate of 20%. There was a final sample of 437 participants. Seven participants reported imagined police interviews or remembered original event details that were not shown in the video and specifically seemed to be generated by an large language model. Specifically, each of these participants generated double the average response length of other participants, described a similar robbery, and mentioned being interviewed by a Detective Johnson. Thus, those seven participants were removed from analysis and the final sample included 430 participants. Participants were 18-35 years old (M = 27.51, SD = 4.91). There was a single participant that reported to be 43 years old. It was unclear if this was a typo (i.e., should be 34) or if they were indeed 43. The participant was retained in the analysis given that even if the participant was 43, this is not old enough to possibly engage in EFT or memory retrieval significantly differently compared to younger adults (e.g., Addis et al., 2008; Gallo et al., 2011). Participants identified as women (n = 226), men (n = 187), non-binary (n = 9), transmen (n = 3), or omitted their response (n = 5). Participants identified as White (n = 218), Black (n = 120), having multiple racial or ethnic identities (n = 29), Latino/Hispanic (n = 26), Southeast Asian (n = 16), East Asian (n = 10), South Asian (n = 8), or Middle Eastern (n = 1), Native American or Native Alaskan (n = 1), Native Hawaiian or Pacific Islander (n = 1). Participants reported their highest levels of education as attending some high school (n = 2),

receiving a high school diploma (*n* = 56), attending some college (*n* = 87), receiving an Associate's degree (*n* = 44), receiving a Bachelor's degree (*n* = 174), or receiving a Master's degree or above (*n* = 67). Most participants reported only being exposed to English at home and at school through the age of 5 (*n* = 354), but some reported being exposed to English and at least one other language through the age of 5 (*n* = 75) and someone reported having no regular English exposure through the age of 5 (*n* = 1).

Materials

All the materials were the same as those used in Experiment 1, except for the police interviews.

Mock Police Interview (Misinformation). The mock interview was a cued-recall memory test where participants were asked specific questions about the original event (Greene et al., 2022). Of note, the questions were edited to be more appropriate for American participants (i.e., changing words to have American spelling and the American word for a baby stroller (e.g., stroller instead of buggy). Additionally, all questions were modified to require memory retrieval of an item, thus removing all yes/no questions. Four of the questions included neutral details that neither confirmed nor contradicted what participants saw in the original event (e.g., "What speed do you estimate the cyclist was moving at when he approached the **traffic** light?"). The other four questions included misleading details that contradicted what participants saw in the original event (e.g., "What speed do you estimate the cyclist was moving at when he approached the **red** light?"). There were two versions of the questions which participants were randomly assigned. These versions were counterbalanced to vary which critical details were misleading or neutral (see Appendix H). Each question appeared on its own webpage and participants could not go

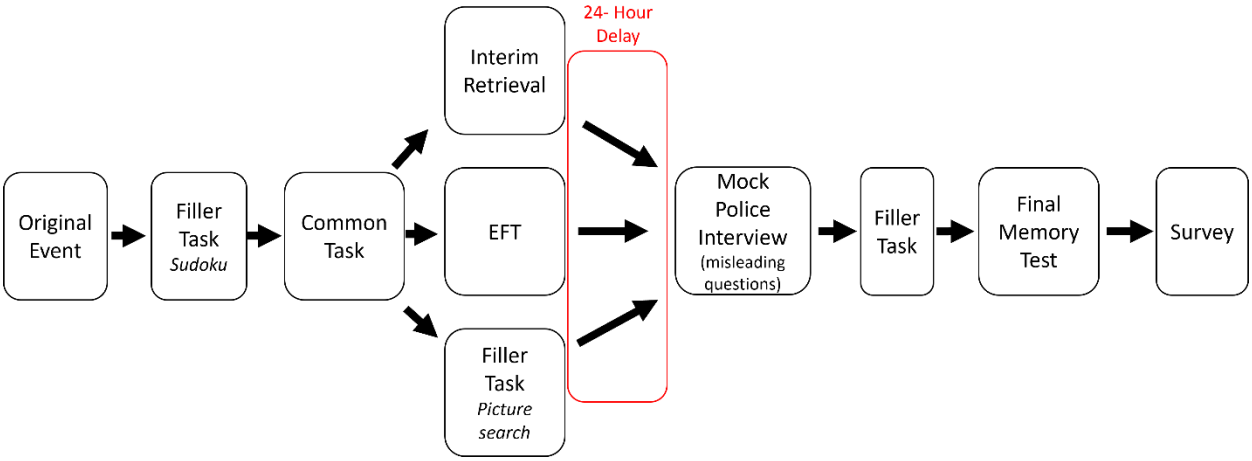
back to previously answered questions. Participants had as much time as they wanted to type out their answer to each question.

Final Memory Test. The final memory test consisted of eight questions relating to the original event (see Appendix H). All the questions had possible answers that were related to the neutral and misleading details that participants were exposed to in the mock police interview. Each question appeared on its own webpage and participants could not go back to previously answered questions. Participants typed out their answer to each question at their own pace.

Procedure

Figure 4

Experiment 2 Paradigm



Participants began the experiment by giving their informed consent. They then watched the original event (see Figure 4). After the original event they completed a filler task which was a game of sudoku for 10 minutes. Following the filler task, participants were assigned to interim task conditions. As in Experiment 1, all participants completed a common event task of imagining a future birthday or remembering a past birthday. After, participants either engaged in interim retrieval of the original event, EFT of a police interview, or completed a filler task. All

participants returned after a 24-hour delay. Upon return, participants had their mock interview about the original event where misleading information was introduced. Following the interview, participants completed 10 minutes of a distractor task which was picture searches. After the distractor task, participants were given the final memory test. Following the final memory test, participants answered demographic questions, as well as a survey relating to their prior experience with eyewitnesses and related media.

Scoring

Scoring of future and past focused details was identical as in Experiment 1. Two research assistants scored 10 participants from the interim retrieval and EFT conditions. Both research assistants were unaware of the hypotheses of the experiment. There was high consistency between the raters for the interim retrieval condition for past details (r = .99), the EFT condition for future details (r = 1) and the EFT condition for past details (r = 1). All participants in the interim retrieval condition reported 0 future oriented details.

Correct details were scored with a 1 and incorrect details scored with a 0 on the final memory test. Misinformation production was scored as a 1 when participants produced misinformation from the leading questions and a 0 if they did not reproduce that misinformation. "I don't know" responses were scored as incorrect given that participants were specifically directed to provide their best guess if they were unsure instead of responding with "I don't know" or "I don't remember". Approximately 1% of questions were given an "I don't know" or "I don't remember" response. A single research assistant scored the cued recall test with instructions on scoring information as correct, incorrect, "I don't know", or misinformation from the initial police interview. These instructions were created prior to data analysis.

Results

All data were deemed suitable for the planned analysis (See Appendix I for complete data screening).

The impact of the interim task on memory of the original event

Intercept only model

I calculated an empty model only including a fixed effect of intercept, and a random effect each for my two nesting variables, question and participant. The empty model was used to assess if significant predictors resulted in a better fitting model for the data compared to a model with no predictors. In the empty model, the fixed intercept was not statistically significant (z = -1.43, p = .153). The random intercept of participant (CI 95% = [0.50, 0.75]) and the random intercept of question (CI 95% = [0.53, 1.47]) were statistically significant. I calculated the ICC and determined that including a random intercept of participant and question accounted for 24.40% of the variation in final memory test accuracy.

The model – question level predictors

I first evaluated the impact of the question level predictor of question type (neutral, misleading) on final memory test accuracy. Neutral questions were associated with significantly higher memory accuracy on the final memory test compared to misleading questions ($\beta = 0.58$, $p < .001$). Including the fixed effect of question type significantly decreased deviance in the model compared to the empty model ($\chi^2 = 55.61$, $p < .001$).

The model – participant level predictors

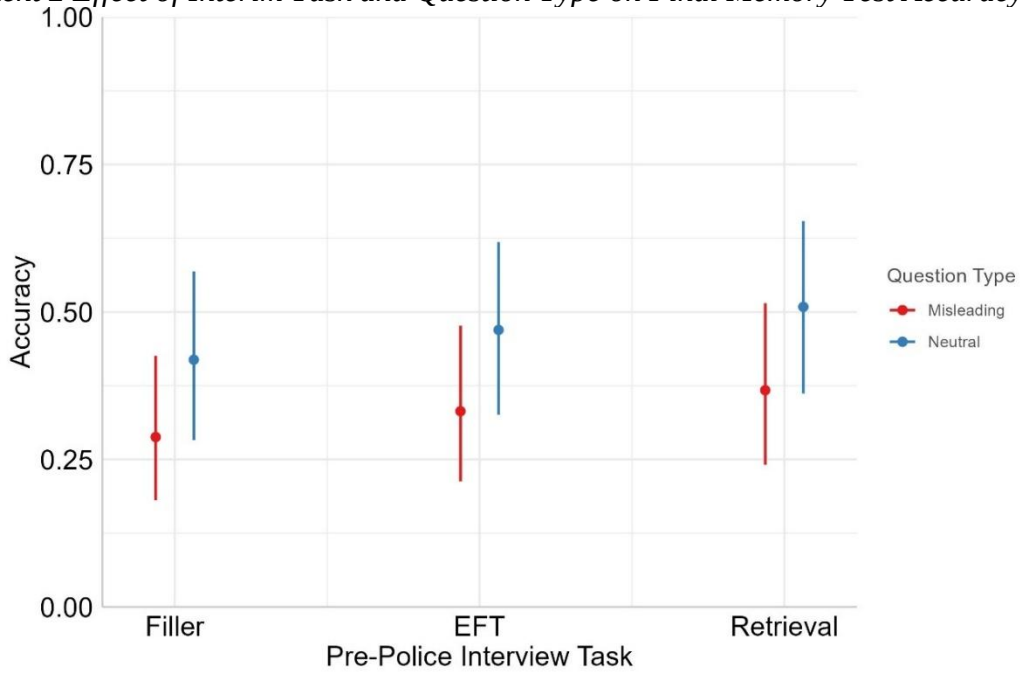
Next, I evaluated if the participant level predictors significantly predicted final test accuracy above question type. The filler condition was the referent within this model. Neutral questions were associated with significantly higher memory accuracy on the test compared to

misleading questions (β = 0.58, $p < .001$). The retrieval condition was associated with significantly higher memory accuracy on the final test compared to the filler condition (β = 0.36, $p = .002$). The EFT condition was not associated with significantly higher final test accuracy compared to the filler condition (β = 0.20, $p = .093$). Including the fixed effect of interim task significantly improved the fit of the model compared to the model that only included the fixed effect of question type ($\chi^2 = 9.21$, $p = .010$).

I estimated a model that included the fixed effect of interim task (EFT, interim retrieval, filler) and the interaction between interim task and question type. There was no significant interaction between interim task and question type for either the EFT (β = -0.06, $p = .770$) or the retrieval condition (β = 0.02, $p = .896$). Thus, I retained the model that included only the fixed effects of question type and interim task without the interaction (See Figure 5).

Figure 5

Experiment 2 Effect of Interim Task and Question Type on Final Memory Test Accuracy



Note. Accuracy is a proportion of correct answers during the final memory test.

The impact of interim task on misinformation production

Intercept only model

I calculated an empty model only including a fixed effect of intercept, and a random effect each for my two nesting variables, question and participant. In the empty model, the fixed intercept was statistically significant (z = -0.72, p = .028). The random intercept of participant (CI 95% = [0.05, 0.45]) and the random intercept of question (CI 95% = [0.60, 1.68]) were statistically significant. I calculated the ICC and determined that including a random intercept of participant and question accounted for 22.40% of the variation in final memory test misinformation production.

The model – question level predictors

I first evaluated the impact of the question level predictor of question type (neutral, misleading) on final test misinformation production. Neutral questions were associated was significantly lower misinformation production compared to misleading questions (β = -1.12, p < .001). Including the fixed effect of question type significantly decreased deviance in the model compared to the empty model (χ^2 = 200.31, p < .001).

The model – participant level predictors

Next, I evaluated if the participant level predictors significantly predicted final test misinformation production above question type. Within the model, the referent was the filler condition. I estimated a model that included the fixed effect of the interim task and question type. There was no significant association between misinformation production and the interim task (EFT: β = 0.01, p = .948; interim retrieval: β = -0.07, p = .482) compared to the filler condition. Thus, the final model retained was the model that included a single predictor variable of question type.

The role of event construction in memory accuracy and misinformation production

The role of event construction on memory accuracy in the EFT condition.

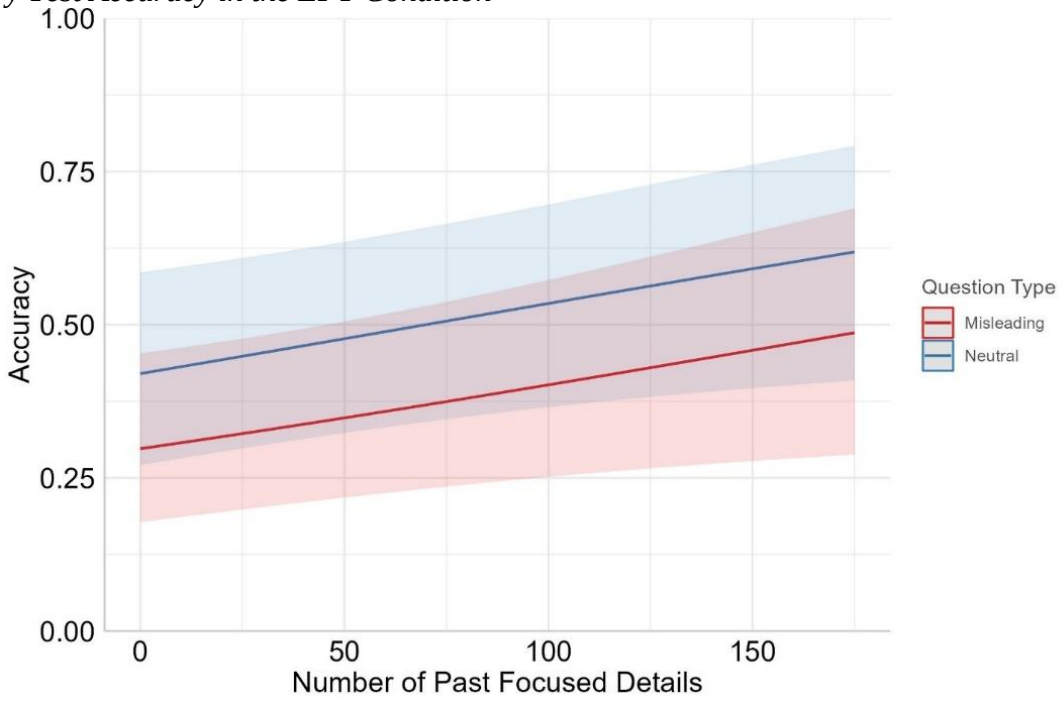
Intercept only model. I calculated an empty model only including a fixed effect of intercept, and a random effect each for my two nesting variables, question and participant. This model only included participants from the EFT condition. In the empty model, the fixed intercept was not statistically significant (z = -0.40, p = .213). The random intercept of participant (CI 95% = [0.25, 0.74]) and the random intercept of question (CI 95% = [0.54, 1.59]) were statistically significant. I calculated the ICC and determined that including a random intercept of participant and question accounted for 23.60% of the variation in final memory test accuracy in the EFT condition.

The model – question level predictors. I first evaluated the impact of the question level predictor of question type (neutral, misleading) on final test accuracy. Neutral questions were associated was significantly higher accuracy compared to misleading questions ($\beta = 0.54$, $p < .001$). Including the fixed effect of question type significantly decreased deviance in the model compared to the empty model ($\chi^2 = 14.71$, $p < .001$).

The model – participant level predictors. Next, I evaluated if the participant level predictors significantly predicted final test accuracy above question type. I first added a fixed estimate of past focused details generated during EFT to a model including the fixed estimate of question type. Neutral questions were associated with significantly higher memory accuracy compared to misleading questions ($\beta = 0.54$, $p < .001$). The number of past focused details was positively associated with memory accuracy ($\beta = 0.01$, $p = .029$; see Figure 6). Including the fixed effect of past focused details significantly decreased deviance in the model compared to the empty model ($\chi^2 = 4.65$, $p = .031$).

Figure 6

Experiment 2 Association Between Number of Past Focused Details and Question Type on Final Memory Test Accuracy in the EFT Condition



Note. Accuracy is a proportion of correct answers during the final memory test.

I then estimated a model that included a fixed effect of future focused details generated during EFT to the above model. The number of future focused details was not significantly associated with memory accuracy for the original event (β = -0.0005, p = .773). Thus, the final model retained included the fixed effect of question type and the fixed effect of number of past focused details.

The role of event construction on misinformation production in the EFT condition.

Intercept only model. I calculated an empty model only including a fixed effect of intercept, and a random effect each for my two nesting variables, question and participant. This model only included participants from the EFT condition. In the empty model, the fixed intercept was statistically significant (z = -0.71, p = .044). The random intercept of participant was not significant (CI 95% = [0.00, 0.58]) but the random intercept of question was significant (CI 95%

= [0.60, 1.78]). I calculated the ICC and determined that including a random intercept of participant and question accounted for 24.1% of the variation in final test misinformation production in the EFT condition. In the following models, I only included a random effect of question type, given that the random effect of participant was insignificant. I retained the random intercept of participant given that the CI included large positive values.

The model – question level predictors. I first evaluated the impact of the question level predictor of question type (neutral, misleading) on misinformation production. Neutral questions were associated was significantly lower misinformation production compared to misleading questions ($\beta = -0.95$, $p < .001$). Including the fixed effect of question type significantly decreased deviance in the model compared to the empty model ($\chi^2 = 44.90$, $p < .001$).

The model – participant level predictors. Next, I then estimated a model that included a fixed effect of past focused details generated during EFT in the model including the fixed effect of question type. The number of past focused details was not significantly associated with misinformation production ($\beta = -0.002$, $p = .225$).

I then estimated a model that included a fixed effect of future focused details generated during EFT in the model including the fixed effect of question type. The number of future focused details was not significantly associated with misinformation production ($\beta = -0.0005$, $p = .740$). Thus, the final model retained included the fixed effect of question type.

The role of event construction on memory accuracy in the interim retrieval condition.

Intercept only model. I calculated an empty model only including a fixed effect of intercept, and a random effect each for my two nesting variables, question and participant. This model only included participants from the interim retrieval condition. In the empty model, the fixed intercept was not statistically significant ($z = -0.25$, $p = .213$). The random intercept of

participant (CI 95% = [0.38, 0.80]) and the random intercept of question (CI 95% = [0.48, 1.41]) were statistically significant. I calculated the ICC and determined that including a random intercept of participant and question accounted for 22.20% of the variation in final memory test accuracy in the interim retrieval condition.

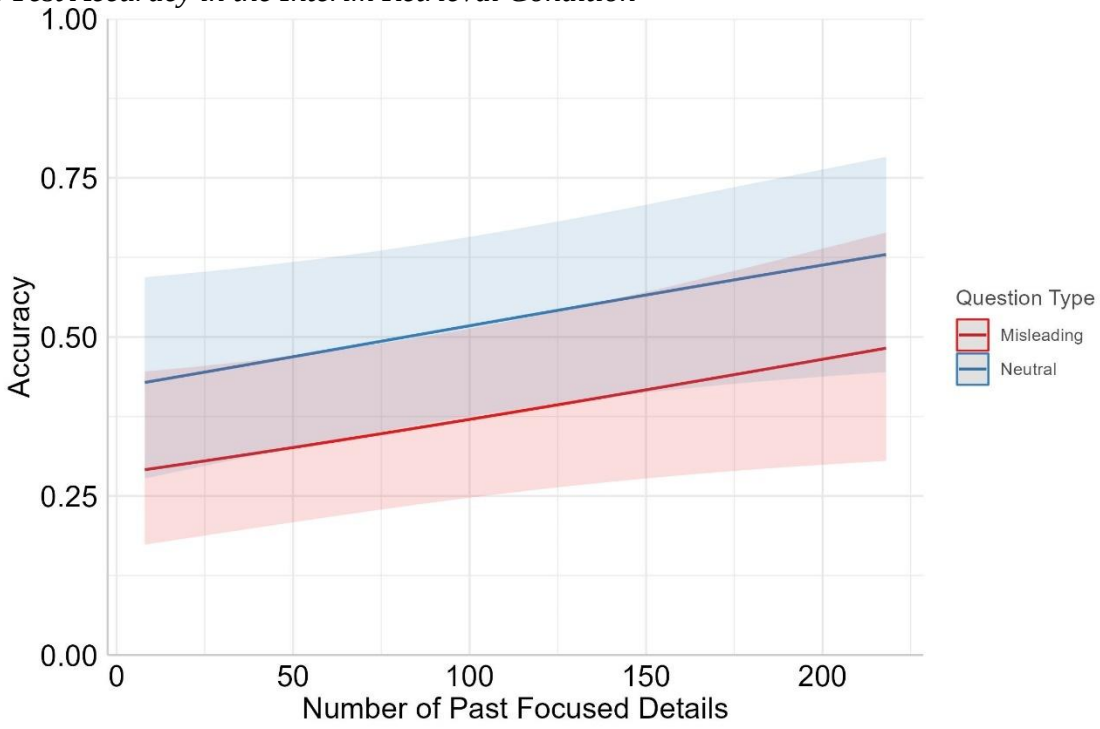
The model – question level predictors. I first evaluated the impact of the question level predictor of question type (neutral, misleading) on final memory test accuracy. Neutral questions were associated was significantly higher accuracy compared to misleading questions ($\beta = 0.60$, $p < .001$). Including the fixed effect of question type significantly decreased deviance in the model compared to the empty model ($\chi^2 = 21.88$, $p < .001$).

The model – participant level predictors. Next, I evaluated if the participant level predictors significantly predicted final test accuracy above question type. I first added a fixed estimate of past focused details generated during initial memory retrieval of the original event to a model including the fixed estimate of question type. Neutral questions were associated with significantly higher memory accuracy compared to misleading questions ($\beta = 0.60$, $p < .001$). The number of past focused details was positively associated with memory accuracy in the interim retrieval condition ($\beta = 0.004$, $p = .047$). Including the fixed effect of past focused details significantly decreased deviance in the model compared to the empty model ($\chi^2 = 3.89$, $p = .048$).

I then investigated if adding an interaction between past focused details and question type further benefited the model. The interaction was not significantly associated with memory accuracy ($\beta = 0.005$, $p = .105$). Thus, the final model retained included a fixed effect of question type and a fixed effect of number of past focused details produced during interim retrieval (See Figure 7).

Figure 7

Experiment 2 Association Between Number of Past Focused Details and Question Type with Final Test Accuracy in the Interim Retrieval Condition



Note. Accuracy is a proportion of correct answers during the final test.

The role of event construction on misinformation production in the interim retrieval condition.

Intercept only model. I calculated an empty model only including a fixed effect of intercept, and a random effect each for my two nesting variables, question and participant. This model only included participants from the interim retrieval condition. In the empty model, the fixed intercept was statistically significant (z = -0.77, p = .024). The random intercept of participant was not significant (CI 95% = [0.00, 0.44]) but the random intercept of question was significant (CI 95% = [0.59, 1.74]). I calculated the ICC and determined that including a random intercept of participant and question accounted for 21.50% of the variation in final memory test misinformation production in the retrieval condition. Given that the random intercept of

participant was not significant, it was removed from the empty model. I retained the random intercept of participant given that the CI contained large positive values.

The model – question level predictors. I first evaluated the impact of the question level predictor of question type (neutral, misleading) on final memory test accuracy. Neutral questions were associated was significantly lower misinformation production compared to misleading questions ($\beta = -1.28$, $p < .001$). Including the fixed effect of question type significantly decreased deviance in the model compared to the empty model ($\chi^2 = 21.88$, $p < .001$).

The model – participant level predictors. Next, I evaluated if the participant level predictors significantly predicted misinformation production in the interim retrieval condition accuracy above question type. I first added a fixed estimate of past focused details generated during interim retrieval of the original event to a model including the fixed estimate of question type. The number of past focused details was not significantly associated with misinformation production ($\beta = 0.001$, $p = .464$). Thus, I retained the model that included the fixed effect of question type.

Phenomenological experience (MCQ) of EFT and memory retrieval

In the following analyses almost all tests are Welch's T-Tests. I also calculated Wilcoxon Summed Ranked Tests due to violations of t-test assumptions. See data screening for more details (Appendix I).

Phenomenological experience (MCQ) of imagining a police interview compared to remembering the original event.

There were no statistically significant differences in participants' phenomenological experience of pre/reliving the events, $t(281.44) = -0.64$, $p = .886$, $d = 0.08$, the emotional intensity of the events, $t(282) = -1.43$, $p = .153$, $d = 0.17$, the events coming to the participant in

words, $t(281.81) = -0.877$, $p = .382$, $d = 0.10$, or the story-like quality of the events, $t(274.82) = 1.49$, $p = 0.137$, $d = -0.18$.

Participants rated their physiological reaction to the imagined police interview as stronger ($M = 3.56$) compared to remembering the original event ($M = 3.14$), $t(273.39) = 3.16$, $p = .002$, $d = 0.38$. Participants rated the valence of the remembered original event as more negative ($M = -0.70$) compared to the imagined police interview ($M = -0.28$), $t(281.99) = -2.68$, $p = .008$, $d = 0.32$. Participants rated the remembered original event as having higher visual clarity ($M = 5.89$) compared to the imagined police interview ($M = 5.63$); $W = 11571$, $p = .021$, $r = 0.14$. Participants also rated the setting as being clearer for the remembered original event ($M = 5.89$) compared to the imagined police interview ($M = 5.51$) $t(241.50) = 2.72$, $p = .007$, $d = 0.36$. Lastly, participants rated their imagined police interview has having higher audio clarity ($M = 5.12$) compared to the remembered original event ($M = 3.10$) $t(261.90) = -9.27$, $p < .001$, $d = 1.09$.

Phenomenological experience (MCQ) of imagining a birthday compared to imagining a police interview.

I compared the phenomenological experience of EFT of a future birthday to the phenomenological experience of EFT of a police interview to determine if there were any perceived differences in imagining a frequently experienced event versus a relatively novel event. There were no differences in participants' phenomenological experience of pre/reliving the events ($t(263.34) = 0.14$, $p = .886$, $d = 0.02$), the emotional intensity of the events ($t(265.26) = 0.71$, $p = .477$, $d = 0.09$), the physiological reaction to the events ($t(264.63) = -0.88$, $p = .076$, $d = -0.22$), the visual clarity of the events ($t(264.42) = 1.75$, $p = .082$, $d = 0.21$), the auditory clarity of the events ($t(263.10) = 0.12$, $p = .905$, $d = 0.01$), the events coming to the participant in words

($t(265.68) = 1.29$, $p = .199$, $d = 0.16$), or the story-like quality of the events ($t(264.37) = 0.71$, $p = 0.479$, $d = 0.09$).

Participants rated the imagined birthday ($M = 1.54$) as being significantly more positive than the imagined police interview ($M = -0.28$; $W = 14754$, $p < .001$, $r = 0.57$). Participants also rated the setting of the imagined birthday as being significantly clearer ($M = 5.98$) compared to the police interview ($M = 5.51$; $W = 10908$, $p = .002$, $r = 0.19$).

Evaluation of sensory versus emotional focus of imagined police interviews

The last analysis that I conducted was an evaluation of sensory and emotional details included when participants engaged in EFT. This analysis was to determine how participants either focused on aperceptual (emotional) or perceptual (sensory) details when imagining an event. Importantly, in Experiment 2 participants were not explicitly told of focus on either type. I used a Wilcoxon rank sum test for this analysis given that the data were not normally distributed. Participants produced more sensory details on average ($M = 89.60$) compared to emotional details ($M = 4.33$) when imagining the police interview, $W = 17956$, $p < .001$, $r = 0.85$.

Experiment 2 Discussion

In Experiment 2, I investigated how EFT impacted the processing of post-event information. I specifically investigated if EFT resulted in a forward effect of testing as seen in prior interim retrieval research (Chan et al., 2009, 2017; LaPaglia & Chan, 2013).

In the final memory test, half the questions were associated with neutral information that was introduced in the mock police interview and half were associated with misleading information that was introduced in the mock police interview. Across all interim task conditions, memory accuracy for neutral questions on the final test were associated with higher memory accuracy compared to misleading questions. Consistent with Experiment 1 findings, participants

that were in the interim retrieval condition had significantly higher final test accuracy compared to the filler condition. Participants in the EFT condition did not have significantly different final test accuracy compared to the filler condition. Importantly, there were no differences across the interim conditions in misinformation production. These results suggest that interim retrieval benefited eyewitness memory accuracy for the original event without increasing susceptibility to misinformation. These results are in line with prior work that found that interim memory retrieval can benefit later memory accuracy for an original event (e.g., Gabbert et al., 2009; LaPaglia & Chan, 2012).

Consistent with Experiment 1, I also found that the number of past details generated while engaging in interim retrieval was significantly positively associated with final test accuracy. Additionally, the number of past details generated during interim retrieval was not significantly associated with misinformation production on the final memory test. Past focused details generated during EFT of the police interview were also significantly positively associated with final test accuracy. These results suggest that if eyewitnesses use crime details to generate their imagined police interview, that their memory of the crime could be improved, even in the face of misinformation exposure. Indeed, in the EFT condition, misinformation production was not found to be significantly associated with the number of past focused details generated during imagining the police interview. Thus, higher rates of including past oriented details did not increase susceptibility to misinformation presented after engaging in EFT. Together, these results of Experiment 2 are consistent with the testing effect literature (for review see Karpicke, 2017) which suggests that retrieval practice can strengthen the memory trace of the retrieved information. The larger theoretical implications of these results are discussed within the general discussion.

In Experiment 2, I further analyzed how reported phenomenological experience of imagining a rarely occurring event (i.e., police interview) compared to imagining a commonly occurring event (i.e., future birthday) and remembering a recently experienced event (i.e., the original event). Consistent with Experiment 1, participants rated their imagined birthday as having a significantly more positive valence compared to their imagined police interviews. Participants also rated the imagined setting of their birthdays as clearer than their imagined police interviews. This may be because people are imagining their birthdays in locations that they have been to before (e.g., their house) resulting in a clearer imagination due to their higher familiarity with the imagined location (e.g., D'Argembeau & Van der Linden, 2012).

When comparing participants phenomenological experiences of engaging in EFT or interim retrieval of the original event, participants reported interim retrieval was associated with higher ratings of visual clarity, and event setting compared to imagining the police interview (EFT). These results are consistent with prior work demonstrating that remembered events were associated with stronger sensory characteristics compared to imagined events (for review see Johnson, 2006). Participants rated the original event as more emotionally negative compared to the imagined police interview. Participants also rated engaging in EFT as resulting in a stronger physiological reaction and being associated with clearer auditory characteristics compared to interim retrieval of the original event. Imagining interacting with police may be more stressful and invoke a stronger bodily response compared to just remembering an event that one had experienced (for review see McCarty, 2016) thus resulting in a stronger phenomenological experience compared to remembering a past event. Further discussion of the theoretical implications of the MCQ results is in the general discussion.

To understand what aspects of an event an eyewitness focuses on when engaging in EFT, I analyzed the production of sensory or emotion focused details during simulation of the event. Participants were found to include significantly more sensory details compared to emotional details when not given instruction on what type of information to focus on. In Experiment 3, I directly manipulated the focus of EFT of a police interview to understand how focus of EFT may affect the impact of EFT on later memory accuracy for the original event.

Experiment 3

EFT can vary in their construction, such as what aspects of an event people choose to simulate as well as how vividly they simulate that event (D'Argembeau et al., 2011; D'Argembeau & Van der Linden, 2004; Rubin, 2014). Research suggests that these variations in EFT can modulate their impact on the experience of the simulation as well as its later memorability. Indeed, familiarity of location and people, the number of times the event is simulated, the temporal distance from the simulated event to when it was simulated, the level of detail simulated, the plausibility, and valence of the event have all been found to impact the experience and later memory for imagined and experienced events (D'Argembeau & Van der Linden, 2012; Gallo et al., 2011; Johnson, 2006; McLelland et al., 2015; Suengas & Johnson, 1988; Thomas et al., 2003). Experiment 3 explored how EFT focused on perceptual details (i.e., sensory details) compared to aperceptual details (i.e., emotions and thoughts) impacted the effect of EFT on eyewitness memory accuracy. Specifically, I explored if modulating the focus of EFT impacted the learning of original event information (backward effect of testing) or the learning of subsequently presented information (forward effect of testing). Additionally, I investigated if the phenomenological experience of EFT is impacted by focusing on perceptual and aperceptual details.

The details that are focused on during EFT can impact the memory characteristics associated with an event. In one study researchers had participants think or talk about experienced or imagined events during the first session (Suengas & Johnson, 1988). In the second session, participants either focused on the perceptual details (i.e., colors, sounds) or apperceptive details (i.e., thoughts, feelings) related to the past or future events they had constructed during the first session. Focusing on the thoughts and feelings decreased the salience of context and sensory characteristics of memories over repeated rehearsal of the event (Suengas & Johnson, 1988). Importantly, this made the memories for experienced and imagined events seem more similar in the subjective number of thoughts and feelings included in the memories. Critically, when the details between experienced and imagined events become more similar, there are higher instances of participants confusing imagined events with experienced events (Johnson, 2006). Thus, I predicted that focusing on aperceptual details would result in more similar phenomenological experiences between the imagined police interview and the imagined future birthday compared to those who focus on sensory details. I further predicted that participants that focused on aperceptual details during EFT would have lower memory accuracy for the original event. Focusing on aperceptual details results in participants retrieving less sensory original event details to construct their EFT. This may have downstream consequences on the impact of EFT on memory accuracy because it will lead to a mismatch between the type of information used to generate the EFT and the type of information on the final memory test. Specifically, the final memory test asked questions relating to sensory details from the original event, and if those details were not retrieved to construct an EFT, then I would not expect to see the benefit of memory retrieval used to simulate an EFT (i.e., backward effect of testing; Karpicke, 2017). This would be due to the mismatch between what is retrieved during the EFT

(aperceptual details such as emotions) and the final memory test (sensory information such as the color of a traffic light). I also hypothesized that lower rates of original event details retrieved during EFT would result in less misinformation production. If participants do not retrieve the details that are later referenced in the misleading police questions because they were focusing on their aperceptual experience, then they should not demonstrate a forward effect of testing (e.g., Gordon et al., 2015)

Method

Participants

The sample size for Experiment 3 was decided by an *a priori* power simulation. I used simr (Green & MacLeod, 2016) to simulate the power for the effect of interim task (interim retrieval, EFT, filler) on memory accuracy for the original event at various sample sizes. A sample of 50 participants was initially taken from Prolific. Ten participants failed attention checks in Experiment 1 and were not invited to complete Experiment 2. Thirty participants completed part two and were included in the power simulation. I used this sample calculate an effect size for the effect of interim task on memory accuracy and used this effect size within the power simulation. A sample size of 270 was needed to power the generalized multi-level model for the effect of interim task (perceptual EFT, aperceptual EFT) with a power of 81.70 (CI 95% [79.16, 84.05]).

An additional 425 participants were sampled from Prolific to complete part 1 of our study. Seventy-four participants failed the attention checks necessary to be invited for session 2 of the study. Three hundred and fifty-one participants were invited to session 2 of the study. Two-hundred and ninety-one participants completed part 2 of the study. During scoring, it was found that an additional twenty participants imagined police interviews that were not related to

the original event video and likely generated from a large language model. These responses were not interpretable and were thus removed from analysis. This resulted in a final sample of 271. Participants were on average 27.91 years old (*SD* = 4.80). The age range of the study was 18-35, but a single participant reported being 47. They were retained in the analysis given that 47 is not considered within the age range of older adults, which could impact one's ability to engage in EFT (e.g., Addis et al., 2008).

Participants identified as women (*n* = 140), men (*n* = 124), agender or non-binary (*n* = 5), and trans or trans men (*n* = 2). Two participants reported their sexual identity instead of their gender identity. Participants identified as White (*n* = 138), having multiple racial or ethnic identities (*n* = 45), Black (*n* = 40), Latino/Hispanic (*n* = 23), East Asian (*n* = 13), Southeast Asian (*n* = 8), South Asian (*n* = 6), Middle Eastern (*n* = 3) or Native American/Alaskan Native (*n* = 1). Participants reported their highest levels of education as attending some high school (*n* = 4), receiving a high school diploma (*n* = 48), attending some college (*n* = 61), receiving an Associate's degree (*n* = 21), receiving a Bachelor's degree (*n* = 111), or receiving a Master's degree or above (*n* = 29). Most participants reported only being exposed to English at home and at school through the age of 5 (*n* = 211). Participants also reported being exposed to English and at least one other language through the age of 5 (*n* = 58) or having no regular English exposure through age 5 (*n* = 2).

Materials

The procedure for Experiment 3 was similar to Experiment 2. The primary difference is the manipulation of perceptual and aperceptual EFT prior to the mock police interview.

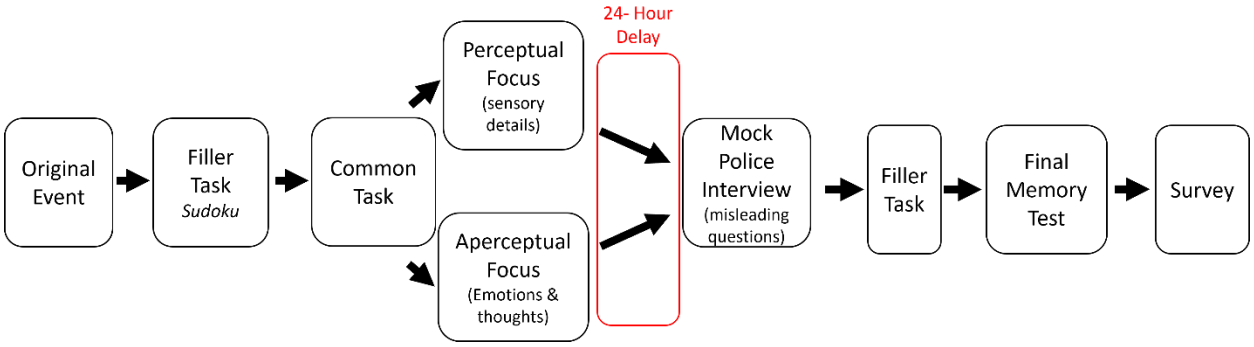
EFT. All participants engaged in EFT. The instructions for this variation of EFT remained mostly the same but differed in what participants were directed to focus on.

Specifically, half of the participants received instructions to focus on the perceptual details (sensory and contextual; see Appendix J) while the other half received instructions to focus on aperceptual details (emotions and thoughts; see Appendix K). Regardless of instructions, participants had 6 minutes to engage in the task and write their imagination.

Procedure

Figure 8

Experiment 3 Paradigm



Participants began the experiment by giving their informed consent. They then watched the original event (See Figure 8). After the original event they completed Sudoku for 10 minutes as a filler task. Following the filler task, all participants engaged in a common event task. Half of the participants (n = 139) focused on perceptual details while engaging in EFT and the other half (n = 132) focused on aperceptual details. All participants were given 6 minutes to write their imaginations. Then, participants rated the experience of imagining the upcoming police interview via the MCQ. Participants returned after a 24-hour retention interval. All participants completed a mock interview that included misleading questions. Following this interview participants completed a picture search task that lasted approximately 10 minutes. After the filler task, participants completed the final memory test. Then, participants completed a survey on their prior experiences with eyewitnesses and filled out the demographic questionnaire. Finally, participants received a compensation code to receive payment for the study.

Scoring

A single research assistant who was unaware of the hypotheses of the experiment coded Experiment 3. Another research assistant also coded 10 participants in the perceptual and 10 participants in the aperceptual condition to check reliability of the scoring methods. Overall, rater reliability in the scoring of the past (r = 0.97) and future details (r = 0.98) in the perceptual condition and the past (r = 0.98) and future (r = 0.96) details in the aperceptual condition were high. For final memory test details, a single research assistant that was unaware of the hypotheses scored the final memory test cued-recall questions for accuracy, misinformation production, and "I don't know" production. On average, participants produced an "I don't know" response on the final memory test on 0.6% of questions (SD = .08).

Results

All data were deemed suitable for the planned analysis (See Appendix M for complete data screening). Of note, certain MCQ analyses were Kruskal Wallis Tests due to ANOVA assumptions being violated.

The impact of EFT focus on memory for the original event

Intercept only model

I calculated an empty model only including a fixed effect of intercept, and a random effect each for my two nesting variables, question and participant. In the empty model, the fixed intercept was not statistically significant (z = -1.38, p = .167). The random intercept of participant (CI 95% = [0.24, 0.59]) and the random intercept of question (CI 95% = [0.49, 1.38]) were statistically significant. I calculated the ICC and determined that including a random intercept of participant and question accounted for 18.80% of the variation in final memory test accuracy.

The model – question level predictors

I first evaluated the impact of the question level predictor of question type (neutral, misleading) on final memory test accuracy. Neutral questions were associated with significantly higher memory accuracy compared to misleading questions ($\beta = 0.77$, $p < .001$). Including the fixed effect of question type significantly decreased deviance in the model compared to the empty model ($\chi^2 = 63.64$, $p < .001$).

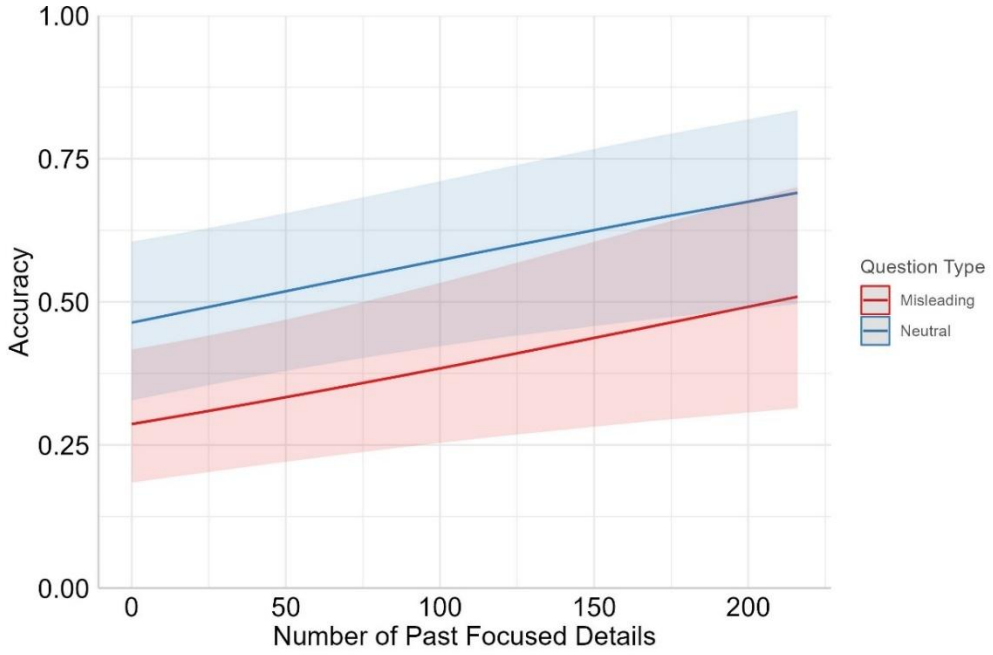
The model – participant level predictors

Next, I evaluated if the participant level predictors significantly predicted final test accuracy above question type. The first predictor that I evaluated was interim task (perceptual EFT, aperceptual EFT). I estimated a model that included the fixed effect of interim task and an interaction term between interim task and question type. Interim task ($\beta = 0.09$, $p = .518$) and the interaction term ($\beta = -0.10$, $p = .617$) did not significantly predict accuracy on the final memory test.

I then included two fixed effects of EFT focus (past focus and future focus). I simultaneously added the two predictors to the model. Question type still predicted final memory test accuracy ($\beta = 0.77$, $p < .001$). The number of past focused details positively predicted memory accuracy on the final memory test ($\beta = 0.004$, $p = .009$). The number of future focused details did not significantly predict final memory test accuracy ($\beta = -0.002$, $p = .055$). Since future focused details did not significantly predict accuracy on the final memory test, I removed the predictor from the model. Including the fixed effect of past focused details significantly improved the fit of the model compared to the model that only included the fixed effect of question type ($\chi^2 = 7.13$, $p = .008$). The final model included the fixed effect of question type and the fixed effect of number of past focused details (See Figure 9).

Figure 9

In Experiment 3 Association of Number of Past Details and Question Type on Final Memory Test Accuracy



Note. Accuracy is a proportion of correct answers during the final memory test.

The impact of EFT focus on misinformation production

Intercept only model

I first estimated an empty model that only included the fixed intercept, a random intercept of participant and a random intercept of question. The fixed intercept was statistically significant (β = -0.65, *p* = .022). The random intercept of question was statistically significant (CI 95% = [0.51, 1.44]) but the random intercept of participant was not statistically significant (CI 95% = [0.00, 0.44]). Together, the random intercepts accounted for 17.40% in variation of misinformation production during the final memory test. The final intercept only model included the fixed intercept and the random intercept of question given that the random intercept of participant was not statistically significant.

The model – question level predictor

I evaluated the impact of the question level predictor of question type on misinformation production on the final memory test. Neutral questions were associated with significantly less misinformation production compared to misleading questions ($\beta = -1.29$, $p < .001$). Including a fixed effect of question type in the model significantly decreased deviance in the model compared to the empty model ($\chi^2 = 168.17$, $p < .001$).

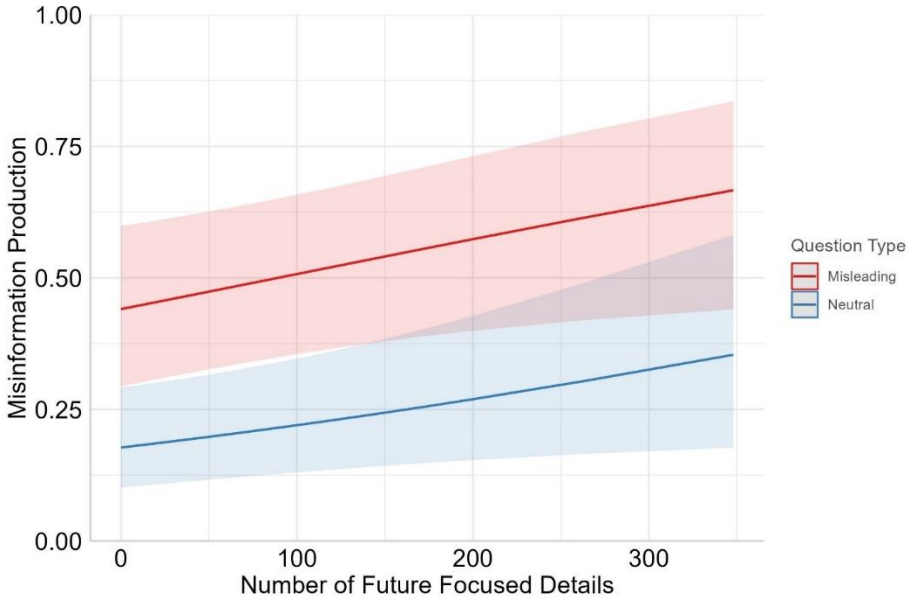
The model – participant level predictors

Next**,** I evaluated the impact of participant level predictors on misinformation production on the final memory test. I first estimated a model that included a fixed effect of interim task and an interaction term between interim task and question type. Interim task ($\beta = -0.19$, $p = .173$) and the interaction term ($\beta = 0.29$, $p = .140$) did not significantly predict misinformation production.

Next, I evaluated the impact of past focused and future focused EFT on misinformation production. The number of future focused details positively predicted misinformation production on the final memory test ($\beta = 0.03$, $p = .036$). Past focused details during EFT did not significantly predict misinformation production ($\beta = -0.002$, $p = .197$). Thus, I removed the past focused predictor. Including a fixed effect of future focus significantly improved the model fit compared to the model that only included a fixed effect of question type ($\chi^2 = 4.55$, $p = .033$). The final model included a fixed effect of question type and a fixed effect of number of future focused details (See Figure 10).

Figure 10

Experiment 3 Association Between Number of Future Details and Question Type on Misinformation Production During the Final Memory Test



Note. Accuracy is a proportion of correct answers during the final memory test.

The role of EFT focus on phenomenological experience (MCQ)

I evaluated if there were any differences in the phenomenological experience of engaging in EFT due to the focus of EFT (perceptual, aperceptual), and event type (birthday, police interview). This analysis allowed me to evaluate if EFT focus impacted the perceived experience of an event, as well as compare if simulated police interviews differed in their perceived experience compared to a highly experienced event (birthday).

EFT focus did not impact participant's ratings of pre-experiencing an event ($F(1,538) = 1.56$, $p = .213$, $\eta_p^2 = 0.003$). There was also no reported difference in the ratings of pre-experiencing the police interview or future birthday during EFT ($F(1,538) = 1.39$, $p = .238$, $\eta_p^2 = 0.003$). EFT focus did not impact participant's ratings that an imagined event was imagined in words, $F(1,539) = 0.01$, $p = .908$, $\eta_p^2 < 0.001$. There were also no differences in the reports of imagining a future birthday and police interview in words ($F(1,539) = 1.61$, $p = .206$, $\eta_p^2 < 0.001$).

There was a significant difference in reported emotional intensity of simulated police interviews ($F(1,538) = 9.38$, $p = .002$, $\eta_p^2 = 0.02$). Specifically, participants in the aperceptual condition reported stronger emotional reactions ($M = 4.90$) compared to participants in the perceptual condition ($M = 4.46$). There was no difference in reported emotional intensity between event types ($F(1,538) = 1.09$, $p = .297$, $\eta_p^2 = 0.003$).

There was a significant difference in reported physiological intensity of imagining an event due to event type (birthday, police interview; $F(1,538) = 4.46$, $p = .035$, $\eta_p^2 = 0.008$). Specifically, participants reported a stronger physiological experience of imagining a police interview ($M = 3.73$) compared to when they imagined their birthday ($M = 3.38$). There was no difference in reported physiological intensity due to EFT focus ($F(1,538) = 1.30$, $p = .255$, $\eta_p^2 = 0.002$).

A Kruskal Wallis Test was conducted, and participants were found to focus on perceptual details reported more positively valenced events ($M = 0.85$) compared to participants who focused on aperceptual details ($M = 0.49$; $\chi^2(1) = 5.14$, $p = .023$, $\eta^2 = 0.01$). Event type also significantly impacted reported valence of the imagined event ($\chi^2(1) = 231.07$, $p < .001$, $\eta^2 = 0.43$). Participants rated their future birthdays as significantly more positive ($M = 1.81$) compared to their imagined police interviews ($M = -0.46$).

Participants who focused on perceptual details when imagining the police interview reported higher ratings of visual intensity of their imagination ($M = 5.79$) compared to participants that focused on aperceptual details ($M = 5.57$; $F(1,539) = 5.12$, $p = .024$, $\eta_p^2 = 0.001$). Further, participants rated the visual intensity of imagining their next birthday ($M = 5.96$) as significantly higher than imagining the police interview ($M = 5.40$; $F(1,539) = 27.96$, $p < .001$, $\eta_p^2 = 0.05$).

A Kruskal Wallis Test was conducted, and participants were found to rate the setting of the imagined birthday has clearer (M = 6.06) compared to the setting of the police interview (M = 5.23; $\chi^2(1) = 55.03$, $p < .001$, $\eta^2 = 0.10$). Participants in the perceptual condition rated the setting of their imagined events as clearer (M = 5.73) compared to participants in the aperceptual condition (M = 5.56; $\chi^2(1) = 4.25$, $p = .039$, $\eta^2 = 0.01$).

Participants rated the auditory intensity of imagining the police interview as higher when they focused on perceptual details (M = 5.22) compared to when they focused on aperceptual details (M = 4.90; $F(1,539) = 5.32$, $p = .021$, $\eta_p^2 = 0.01$). Furthermore, participants rated the auditory intensity of their imagined birthdays (M = 5.30) as higher than their imagined police interviews (M = 4.83; $F(1,539) = 12.03$, $p = .001$, $\eta_p^2 = 0.02$).

Lastly, participants reported that their imagined birthdays were more strongly imagined as a story (M = 5.73) compared to their imagined police interviews (M = 5.41; $F(1,539) = 8.10$, $p = .005$, $\eta_p^2 = 0.01$). There was no difference in rating of the perceived story-like quality of events for participants in the perceptual and aperceptual focus conditions ($F(1,539) = 1.75$, $p = .187$, $\eta_p^2 = 0.003$).

Experiment 3 Discussion

Experiment 3 investigated how focusing on perceptual and aperceptual details during EFT impacted the effect of EFT on original event memory accuracy. Focusing on perceptual or aperceptual details during EFT did not significantly influence memory accuracy. Consistent with Experiment 2, final memory test accuracy for neutral questions was higher than misleading questions. Additionally, the number of past focused details generated during EFT positively predicted memory accuracy during the final memory test, while the number of future focused details generated during EFT did not predict memory accuracy during the final memory test.

When evaluating misinformation production on the final memory test, EFT focus (perceptual, aperceptual) did not predict misinformation production. These results suggest that focusing on perceptual or aperceptual details does not influence susceptibility to post-event information. Consistent with Experiment 2, misleading questions suggested during the mock police interview were associated with higher rates of misinformation production on the final memory test compared to neutral questions. Additionally, the number of past focused details generated during EFT did not predict misinformation production on the final memory test. One novel finding in Experiment 3 was that the number of future focused details produced during EFT positively predicted misinformation production on the final memory test. These results suggest that focusing on future oriented details during EFT impacted how participants processed post-event misinformation. Importantly, since past focused details were not predictive of misinformation production, these results do not suggest a retrieval-based explanation of EFT (e.g., Gordon & Thomas, 2014). Instead, future focus is capturing another process that participants are undergoing when imagining a police interview that is then impacting misinformation production. In the general discussion I discuss the various reasons why the number of future oriented details generated while imagining the police interview may impact misinformation susceptibility.

In Experiment 3, I also explored how EFT focus impacted the phenomenological experience of engaging in EFT. In line with previous work, when participants focused on perceptual details when imagining an event, they rated their experience higher in perceptual details compared to those who focused on aperceptual details (Suengas & Johnson, 1988). Specifically, participants who focused on perceptual details rated the visual intensity and clarity of setting of the event as stronger than those who focused on aperceptual details. Participants in

the aperceptual condition rated having a stronger emotional experience while imagining the event compared to participants in the perceptual condition. These findings suggest that when participants were instructed to focus on perceptual or aperceptual aspects of an event, it shaped their phenomenological experience of imagining the event.

Participants that focused on perceptual details rated their imagined events as significantly more positive than participants that focused on aperceptual details. This may be the case given that when focusing on aperceptual details (emotions and thoughts) participants may reflect on a variety of positive and negative emotional experiences, while when participants focus on perceptual details they make less emotional evaluations in the moment. Then, when they are asked to reflect on the event, participants that had a more varied emotional experience may rate the event as less positive than those who just focused on sensory details.

I also evaluated if there were any differences in participant's phenomenological experience of imagining a future birthday compared to the future police interview. Consistent with Experiment 2, participants rated the physiological response to imagining a police interview as more intense compared to imagining their birthday. Additionally, participants rated their imagined birthday as more visually clear, having a clearer setting, having stronger auditory information, and being imagined more like a story compared to their imagined police interviews. These results are consistent with past research that EFT draws from past experiences and imaginations can be richer in detail based on the ability to use prior experience to construct the simulation (Addis et al., 2009; Irish et al., 2012; Szpunar & McDermott, 2008). The implications of these results are discussed fully in the general discussion.

General Discussion

In three experiments, I investigated how participants engaged in EFT of a novel event and how engaging in EFT impacted memory for the original event. I specifically explored if engaging in EFT impacted how participants learned previously presented information (backward effect of testing) or subsequently presented information (forward effect of testing). I posit that the observed relationship between past detail generation during EFT and memory accuracy for the original event are indicative of a backward effect of testing, similar to what was observed in the interim testing condition (Experiment 1 & 2; e.g., Karpicke, 2017). Across these three experiments I also explored how participants engaged in EFT, and how this experience (e.g., MCQ) of imagining a novel event varied from remembering or imagining events one had experienced before. I posit that the results across the three experiments suggest that participants used semantic based memories to generate EFT that may be comparably rich in detail to EFT that were generated with episodic memories.

Below, I propose how the results of each of the three experiments, as well as prior literature, support these findings. Additionally, I discuss future research directions that would further clarify our understanding of the present results.

The testing effect of EFT

The testing effect is the finding that interim memory retrieval results in better memory for retrieved details over time (for review see Karpicke, 2017). In Experiment 1 & 2, there was strong evidence of a testing effect for participants in the interim retrieval condition as their memory accuracy on the final test was significantly higher compared to the filler condition. Additionally, there was a positive relationship between the number of past focused details participants generated during interim retrieval and final test accuracy. While engaging in EFT was not associated with significantly different final test accuracy compared to the filler condition

(Experiment 1 & 2), there was a significant positive association between the number of past focused details generated during EFT and final memory accuracy (Experiment 2 & 3). These results suggest that memory retrieval that occurs to simulate an imagined event can impact later memory accuracy for an original event. In other words, if eyewitnesses use their memory of the crime to imagine how the police might interview them, this process of imagining an event may impact memory for the crime.

Engaging in memory retrieval, either in the interim retrieval condition or EFT condition, was not found to impact processing of misleading post-event information in the three experiments. Neither past (Experiment 2 & 3) nor future (Experiment 2) focused details generated during interim retrieval or EFT were significantly associated with misinformation production. One explanation for why retrieval of original event details benefitted memory accuracy while not impacting misinformation production may be due to changes in modality of presentation. Prior work found that when misleading information was presented in a narrative format RES occurred, yet when it was presented via misleading questions, there was a reduced susceptibility to misinformation (LaPaglia & Chan, 2013). The researchers suggested that when misinformation was presented as a question, the act of attempting to answer the question brought other related information from memory to mind resulting in participants noticing the inconsistency. Thus, in the present experiments, participants may detect misleading details through retrieving original event details to answer misleading questions. This error detection could lead participants to label this misleading information as misinformation during encoding. Additionally, through retrieving original event details to answer these questions, participants may benefit from the retrieval attempt (e.g., Karpicke, 2017). During the final memory test,

participants might be able to use their memory of the misinformation, as well as their memory for original event details, to correctly answer memory test questions.

Interim retrieval resulted in a backward effect of testing. That is, participants who engaged in interim retrieval demonstrated better final memory performance than participants who engaged in an interim filler task. Importantly, interim retrieval did not result in RES. RES may not have emerged because the interim tasks were a free recall test in (Experiment 2) and an open imagination task (Experiment 2 & 3). Free recall or undirected imagination may result in participants not retrieving critical original event details that are later manipulated. Prior work suggests that retrieval of critical original details may be a necessary condition of RES. Research suggests that when original details are retrieved, participants spend more time processing those modified details in a later post-event narrative, than when original details are not initially retrieved (Gordon et al., 2015). Importantly, more time processing misleading details was directly associated with RES. In the present study, if participants did not retrieve the original event details that were related to the later presented misinformation during the interim task, then they may be less susceptible to RES.

Past research has found that initial free recall can result in RES (Wilford et al., 2014). Yet, in those experiments participants were given over 20 minutes to remember the original event while participants in the present experiments were only given 6 minutes to engage in their imagination/retrieval tasks. Less time to engage in the task likely resulted in participants retrieving less information from the original event. Retrieving less information could also result in a smaller likelihood that the information that was retrieved overlapped with the post-event information. In the present experiments, initial free recall of events or free simulation of future

events was implemented to get as close to reality of what an eyewitness would do in the real world when spontaneously remembering the crime or imagining a police interview.

The number of future focused details used to generate a future police interview positively predicted misinformation produced on the final memory test. Specifically, in Experiment 3 I found that the number of future focused details generated during EFT positively predicted misinformation production on the final test. This was an unexpected result, given that I originally hypothesized that retrieval-based mechanisms (i.e., forward effect of testing) would impact susceptibility to misinformation and future focused details did not rely on the retrieval of past focused details. One explanation for these results is that imagining the police interview increased the perceptual fluency of the experienced mock police interview. Perceptual fluency is the ease with which information is processed due to prior exposure (Jacoby & Dallas, 1981). The act of imagining a detailed police interview may result in participants more easily processing the information they encounter during the mock police interview because it is consistent with the scenario they imagined during EFT. Importantly, fluency plays a role in our ability to recognize errors in statements. Indeed, prior work found that when information was more difficult to read due to font size, participants were more likely to notice an error in a statement than if the text was easier or more fluent to process (Song & Schwarz, 2008). In the present experiments, the increased processing fluency of mock interview questions due to imagining the police interview may have reduced the likelihood of detecting misinformation when presented. Later, on the final test, participants may have been more likely to produce those misleading details because they failed to encode them as discrepant and their more recent exposure to this information made it easier to retrieve (i.e., retrieval fluency). Prior research suggested that participants may produce the more recently encountered post-event information on later memory tests because they

terminate a more effortful memory search once the quick to retrieve answer is obtained (Thomas et al., 2010).

Future focused details generated during EFT may further be capturing an aspect of individual differences in visualization capabilities. Prior research found that visual-spatial processing abilities were related to the number of sensory descriptions reported when imagining specific future events (D'Argembeau et al., 2010). Thus, when the current experiments measure future focused detail production they may be measuring an aspect of individual differences in visualization capabilities. This is important to consider given that higher visualization capabilities have also been found to result in higher rates of imagination inflation (e.g., Horselenberg et al., 2000; Paddock et al., 1998) and misinformation susceptibility (Tomes & Katz, 1997). Thus, future work in this area should directly measure visualization capabilities to see if this individual difference may account for susceptibility to misinformation possibly above and beyond the number of future focused details a participant produces when engaging in EFT.

The relationship between future focused details and misinformation production should be interpreted with caution given that the relationship was only observed in Experiment 3. One reason for these results may be the differences in EFT instructions between Experiment 2 and 3. Specifically, in Experiment 2, participants were asked to freely imagine the police interview and not given instructions on what sensory or emotional details to focus on. When evaluating the contents of these imaginations, participants generated significantly more perceptual (sensory) details when imagining the police interview compared to aperceptual (emotional) details. In Experiment 3, participants were directed to either focus on perceptual or aperceptual details. Thus, when looking across the two conditions there were significantly more people focusing on the aperceptual aspects of imagining a police interview compared to those in the EFT condition

in Experiment 2. Importantly, focusing on emotional information during event simulation has been found to impact source monitoring capabilities (for review see Johnson, 2006). That said, when comparing the effect of future focused detail generation on misinformation production there were no differences in the strength of the association between the perceptual and aperceptual conditions in Experiment 3. Thus, the emotional focus on the event may not underlie the association between future focused details and misinformation production. Further research should endeavor to replicate these findings across a variety of contexts and samples to better understand the stability and mechanisms for this observed effect.

The phenomenological experience of EFT

Prior research suggests that people should have a more detail rich remembering/imagining experience when they have prior experiences to scaffold their imagined/remembered events (for review see Schacter et al., 2017). Thus, I hypothesized that participants should have a stronger recollective experience for the original event compared to the imagined police interview because participants can draw from the recent experience of the original event while they likely have not been interviewed by police before. Consistent with this hypothesis, in both Experiment 1 and 2 participants rated their phenomenological experience of interim retrieval (remembering the original event) as stronger in sensory ratings (e.g., visual clarity) as compared to imagining a police interview (EFT of a novel event).

I also expected that imagined novel events (i.e., police interview) would be associated with weaker sensory experiences as compared to imagined experienced events (e.g., future birthday). This would be due to the lack of prior episodic memories associated with the novel event compared to the prior experienced event resulting in weaker phenomenological experience of EFT (Arnold et al., 2011; D'Argembeau & Van der Linden, 2012; Schacter et al., 2017).

Interestingly, I did not consistently find these differences across experiments. Indeed, the only consistent phenomenological difference between imagined police interviews and birthdays was their rated valence. In Experiment 2, participants rated their imagined birthday as having a clearer setting compared to the imagined police interview. Further in Experiment 3, they also rated their imagined birthdays as having stronger visual details, stronger auditory details, and being experienced more like a story compared to their imagined police interviews.

The similarity in MCQ ratings between an imagined police interview and an imagined upcoming birthday have interesting implications for our theoretical understanding of how future events are constructed. Indeed, the stronger sensory experience of imagining a future birthday compared to a police interview in Experiment 3 is consistent with prior research (Arnold et al., 2011; D'Argembeau & Van der Linden, 2012; Johnson, 2006). These results suggest that participants may draw from episodic memories related to their previous birthdays to generate their imagined birthday and rely more on semantic memories associated with general knowledge of police interviews to engage in EFT of the upcoming police interview. Yet, these differences are only observed in Experiment 3. In smaller samples sizes, participants only rated differences in valence (Experiment 1 & 2), and setting clarity (Experiment 2) between the imagined police interview and imagined future birthday. These results suggest that even when imagining an EFT for this novel event, participants seemingly are able to generate an EFT that is almost as rich in sensory details as an EFT that is associated with a wealth of prior experiences. Thus, participants may be able to use semantic memories to generate a novel EFT that results in a phenomenological experience that seems as robust as an EFT based on experienced events. Future research should further explore the role of semantic memories in novel EFT construction through including older adults in this paradigm. Past research has found that older adults tend to

rely on semantic memory as compared to episodic memory when engaging in EFT compared to younger adults (Schacter et al., 2013). Thus, comparing younger and older adults in this eyewitness EFT task will allow us to better understand if any differences arise in EFT construction that may be due to differences in using semantic and episodic memories to simulate future events.

Participants may have episodic memories of specific pieces of eyewitness related media that they are using to generate the imagined police interview. That said, prior exposure to police media (e.g., movies, TV shows, YouTube videos, podcasts, news articles) was not found to predict memory accuracy, nor perceived clarity of several phenomenological experiences reported on the MCQ. In addition, experimental results were stable when participants who reported having witnessed a crime before were removed. Of note, the current measurements of prior experience may not be accurate enough to evaluate the relationship between prior experience and EFT ability. For example, when asked to report how many TV shows one had previously seen some participants named specific TV franchises while others gave a number. Thus, it is impossible to compare if those who reported TV franchises had seen a similar number of episodes as the person who reported the exact number of episodes they had seen. Additionally, the question relating to experiencing a prior crime was designed to assess if people had prior interactions with police. Yet, given the wording of the question, participants may have reported seeing a crime before but have never been interviewed by police. Thus, they may not have robust knowledge about how the interview process is conducted. Future research should focus on creating comprehensive questions that measure prior knowledge and experience with police interviewing to better understand the relationship between prior experience and EFT capabilities.

One limitation of the present analyses when comparing participants' phenomenological experience of imagining a police interview to a birthday is the inability to account for temporal differences between the events. In Experiment 1, there was a difference in the time given to participants to imagine their upcoming birthday compared to their upcoming police interview. This could have resulted in participants reporting a less intense sensory experience of imagining a future birthday because they had less time to engage in the task. Experiment 2 removed this possible confound through increasing the time given to participants to imagine their future birthday. While this temporal confound was able to be addressed across experiments, I was not able to address the differences *when* in time this imagined event would occur. For all participants, the imagined police interview was happening "tomorrow", but for each participant their future birthday is going to vary in time. For example, one participant could have a birthday a week from now while another could have a birthday 8 months from now. This temporal difference in when the imagined event will occur likely impacts the experience of imagining the event. Indeed, prior work found that events that are further away in time were imagined with less sensory and contextual details compared to events happening sooner (e.g., D'Argembeau & Van der Linden, 2004). Thus, future research is needed to understand if imagined police interviews were rated with similar clarity as imagined birthdays because they were happening "tomorrow" while a participants next birthday could happen at various time points in the future.

Conclusion

Imagining an upcoming police interview can be a self-generated event that impacts eyewitness memory. The contents of the imagined police interview (e.g., the number of past focused details) is predictive of eyewitness memory accuracy. Consistent with testing effect literature, the number of past focused details included in a simulated police interview positively

predicts memory accuracy for the original event. Additionally, there is some evidence to suggest that the number of future focused details included in a simulated police interview positively predicted misinformation production. These results suggest that the act of imagining a police interview may result in a forward effect of testing, such that EFT impacts how participants engage with and later remember post-event information. When examining the phenomenological experience of EFT, participants simulated detail-rich imagined police interviews. These results suggest that participants may use semantic information to simulate imagined events that are phenomenologically similar to an imagined event that one has experienced many times before. Overall, these experiments suggest that EFT can impact eyewitness memory for the crime and that the EFT process is a phenomenologically rich experience.

Appendix A

Experiment 1 Attention Checks

EFT Condition

You will now **imagine** the police interview tomorrow. There are **TWO** main components of imagining the police interview that are important for you as a witness. **First**, and similar to how you imagined your next birthday, you should **imagine** details like **where you will be**, **who will be there**, what the **order of events will be**, and how you **will feel during the police interview**. **Second**, you should also **imagine how you will answer questions about the crime and the details from the crime you will report.**

When imagining the police interview your imagination needs to include **BOTH** components listed above.

The 6-minute timer will start immediately when you advance to the next page. Please advance to the next page when you are ready to start **imagining the police interview.**

1. Based on the instructions you just read above, what is the first component for you to imagine about the police interview?

   Please re-read the instructions above if you are not sure.

   A. The details of the police interview. For example, who would be at the interview

   B. A description of the thief. For example, what the thief was wearing

   C. Not Specified

2. Based on the instructions you just read above, what is the second component for you to imagine about the police interview?

   Please re-read the instructions above if you are not sure.

   A. The questions that police would ask about the crime and the crime details you would respond with

   B. The details about what would happen after the police interview in the future

   C. Not Specified

3. Based on the instructions you just read above, what components should you include in your written imagination of the police interview?

Please re-read the instructions above if you are not sure.

A. The first component

B. The second component

C. Both components

Interim retrieval condition

You will now **remember** the crime that you witnessed earlier. There are **TWO** main components of remembering the crime that are important for you as a witness. **First**, and **similar to how you remembered your last birthday**, you should **remember** details like where you **were**, who **was** there, what the order of events **were**, and how you felt **while the** crime **occurred**. **Second**, you should also **remember details that you think were important about the crime. For example, what was the nature of the crime.**

When remembering the crime your report needs to include **BOTH** components listed above. **The 6-minute timer will start immediately when you advance to the next page**. Please advance to the next page when you are ready to start **remembering the crime.**

1. Based on the instructions you just read above, what is the first component for you to imagine about the police interview?

   Please re-read the instructions above if you are not sure.

A. The details of the police interview. For example, who would be at the interview

B. Imagining what it is going to be like to be interviewed by police

C. Not Specified

2. Based on the instructions you just read above, what is the second component for you to imagine about the police interview?

   Please re-read the instructions above if you are not sure.

   A. The details about what would happen after the police interview in the future

B. The details that seemed important about the crime

C. Not Specified

3. Based on the instructions you just read above, what components should you include in your written imagination of the police interview?

   Please re-read the instructions above if you are not sure.

   A. The first component

   B. The second component

   C. Both components

Appendix B

Experiment 1 Memory Retrieval Common Event Instructions

You will now be asked to remember your last birthday. Please remember your last birthday in as much detail as possible. For example, you may remember details like where you were, who was there, the order of events and how you felt during your birthday. You will have 3 minutes to write about this event as you remember it. Please be as precise as possible when remembering your birthday.

If you finish writing before the 3 minutes are up, please go back and continue to add as much information as possible, either relating to what you already wrote or what you may not have written about yet. Remembering your last birthday in as much detail as possible should allow you to truly mentally re-experience the event. After the 3 minutes are up, the page will automatically advance to the next page. You will then answer questions about what it was like to remember your birthday.

When you are ready to start remembering your last birthday, please advance to the next page.

Experiment 1 EFT Common Event Instructions

You will now be asked to imagine your next birthday. Please imagine your next birthday in as much detail as possible. For example, you may imagine details like where you will be, who will be there, what the order of events will be, and how you will feel during your birthday. You will have 3 minutes to write about this event as you imagine it. Please be as precise as possible when imagining your next birthday.

If you finish writing before the 3 minutes are up, please go back and continue to add as much information as possible, either relating to what you already wrote or what you may not have written about yet. Imagining your next birthday in as much detail as possible should allow you to truly mentally pre-experience the event. After the 3 minutes are up, the page will automatically advance to the next page. You will then answer questions about what it was like to imagine your next birthday.

When you are ready to start imagining your next birthday, please advance to the next page.

Appendix C

EFT Condition Instructions

Since you are an eyewitness to the crime, you have been contacted by police to **be interviewed about what you saw.** Today, you will write about **what you think will happen during the police interview tomorrow. Please imagine what giving your statement to police tomorrow will be like** in as much detail as possible. You will have 6 minutes to write about this event as you **imagine** it. Please be as precise as possible when **imagining the interview**.

Imagining the police interview should be a similar process to how you just imagined what your next birthday will be like. For example, you may **imagine** details like where you **will be**, who **will be there**, what the order of events **will be**, and how you **will** feel **during the interview**. **In addition to these details**, you should also **imagine how you will answer questions about the crime and the details from the crime you will report.**

If you finish writing before the 6 minutes are up, please go back and continue to add as much information as possible, either relating to what you already wrote or what you may not have written about yet. **Imagining the interview** in as much detail as possible should allow you to truly mentally **pre-experience** the event. After the 6 minutes are up, the page will automatically advance to the next page. You will then answer questions about what it was like **to imagine the interview.**

----- PAGE BREAK -----

Points to focus on while imagining:

1. Details about the future police interview (for example: who will be there, what will happen

and what will be the order of events, where will you be interviewed, how you will feel while being interviewed)

2. Questions that the police will ask you about the crime you saw at the beginning of this experiment and how you would respond to those questions with what you remember from the crime

Write your imagined interview below.

If you finish writing before the timer is up, please go back and continue to add as much information as possible either relating to what you already wrote or what you may not have written about yet. *Please reference the list of details to focus on above if you are unsure what to write and include both components in your response.*

Appendix D

Interim retrieval condition instructions

Since you are an eyewitness to a crime, you have been contacted by police to be interviewed about what you saw. Today, you will write about what you remember about the crime. Please remember the crime you saw at the beginning of the experiment in as much detail as possible. You will have 6 minutes to write about this event as you remember it. Please be as precise as possible when remembering the crime.

Remembering the crime should be a similar process to how you just remembered what your last birthday was like. For example, you may remember details like where you were, who was there, what the order of events was, and how you felt while the crime occurred. In addition to these details, you should also remember details that you think were important about the crime.

If you finish writing before the 6 minutes are up, please go back and continue to add as much information as possible, either relating to what you already wrote or what you may not have written about yet. Remembering the crime in as much detail as possible should allow you to truly mentally re-experience the event. After the 6 minutes are up, the page will automatically advance to the next page. You will then answer questions about what it was like to remember the crime.

----- PAGE BREAK -----

Points to focus on while remembering:

1. Details about the crime you saw at the beginning of this experiment (for example: who was there, what happened and what were the order of events, where did the crime take place, how

you felt while watching the crime)

2. Details you found important

Write what you remember from the crime below.

If you finish writing before the timer is up, please go back and continue to add as much information as possible either relating to what you already wrote or what you may not have written about yet. *Please reference the list of details to focus on above if you are unsure what to write and include both components in your response.*

Appendix E

Experiment 1 Final Memory Test Instructions

You will now write a statement about what you witnessed during the crime. Please be as precise as possible when remembering the crime. For example, you may remember details like where you were, who was there, what the order of events was, and how you felt while the crime occurred. If you finish writing before the 6 minutes are up, please go back and continue to add as much information as possible, either relating to what you already wrote or what you may not have written about yet. Remembering the crime in as much detail as possible should allow you to truly mentally re-experience the event. After 6 minutes are up, the page will automatically advance to the next webpage. You will then answer questions about what it was like to remember the crime.

----- PAGE BREAK -----

You will now **remember** the crime. Please remember the events in as much detail as possible. Specifically, when and where the event took place as well as who was there, and what happened.

The 6-minute timer will start immediately when you advance to the next page. Please advance to the next page when you are ready to start **remembering the crime.**

----- PAGE BREAK -----

Below write what you remember about the crime you saw during part 1 yesterday.

If you finish writing before the timer is up, please go back and continue to add as much information as possible, either relating to what you already wrote or what you may not have written about yet.

Appendix F

Modified MCQ

1. While remembering/imagining the event, I feel as though I am reliving/pre-experiencing the event.

1	2	3	4	5	6	7
Not at all						As if it were happening now

2. While remembering/imagining, the emotions that I feel are intense.

1	2	3	4	5	6	7
Not at all						Extremely strongly

3. While remembering/imagining the event, I have a physical reaction (I feel tense, sweaty, felt cramps or butterflies in my stomach, my heart pounded, etc.).

1	2	3	4	5	6	7
Not at all						Extremely strongly

4. While remembering/imagining, the emotions are negative or positive.

-3	-2	-1	0	1	2	3
Extremely negative						Extremely strongly

5. While remembering/imagining the event, I can see it in my mind.

1	2	3	4	5	6	7
Not at all						Extremely strongly

6. While remembering/imagining the event, I know the setting.

1	2	3	4	5	6	7
Not at all						Extremely strongly

7. While remembering/imagining the event, I can hear it in my mind.

1	2	3	4	5	6	7
Not at all						Extremely strongly

8. While remembering/imagining the event, it comes to me in words.

1	2	3	4	5	6	7
Not at all						Extremely strongly

9. While remembering/imagining the event, it comes to me in words or in pictures as a coherent story.

1	2	3	4	5	6	7
Not at all						Extremely strongly

Appendix G

Experiment 1 Data Screening

Prior to data analysis, the data were evaluated to ensure that they were appropriate for each planned analysis. Given that participants were randomly assigned to conditions we assume that the data meets our assumptions of independence required for all planned statistical tests.

To evaluate if the data were appropriate for linear regression, I first checked the distribution of the outcome variable of memory accuracy. This evaluation of the outcome variable approximates the distribution of the model residuals which can only be obtained after the model is estimated. The distribution of the outcome variable was evaluated using histogram of the distribution as well as the skew and kurtosis of the variable. The histogram looked approximately normal, and this was supported by the skew < |1| (skew = 0.77) and kurtosis < |3| (kurtosis = -0.35) of the data (Field, 2018). There were three linear regressions planned for Experiment 1. The first only contained categorical variables to assess if there were differences in memory accuracy for the original event across interim task groups (interim retrieval, EFT, filler). To further assess the suitability of the data for this model, I evaluated the homoscedasticity of the outcome variable by interim task group (interim retrieval, EFT, filler). I created a box plot to evaluate if there were comparable amounts of variance in the outcome variable (correct details reported about the original event) across the three interim task conditions. The variance between the groups was approximately similar to each other.

Another linear regression included two continuous predictor variables of number of past and future focused details produced during EFT. To ensure that these two variables were not problematically multicollinear, I calculated a correlation between the two items (r = -0.21).

These items were weakly and negatively correlated with one another, suggesting that they were not multicollinear.

I then evaluated the MCQ data to determine if they were appropriate for one-way analyses of variance (ANOVAs) and Welches T-tests. I evaluated the MCQ responses for the final memory test, the interim retrieval condition, the EFT condition for their birthday and police interview. All MCQ questions had kurtosis < |3| (See Table 9-12). Only four MCQ questions across all four sets of data had a skew > |1| (See Table 1-4). When assessing the qqplots for each MCQ question to evaluate normality, none of the questions for any of the comparisons visually displayed non-normality. For the ANOVA analysis, I then evaluated the variance within each of the initial task conditions by MCQ question through box plots. Any plots that looked as if they were possibly invariant were checked with a Bartlett Test. All MCQ questions were evaluated to be homoscedastic.

Overall, I concluded that the data was appropriate for the planned linear regression, ANOVA analyses, and t-test analyses. Welches t-tests were used given that they do not require assumptions of variation.

Table 9

Experiment 1 Final Memory Test MCQ Ratings Descriptive Statistics

MCQ Question	Mean	*SD*	*n*	Min	Max	Skew	Kurtosis
Reliving	4.79	1.45	131	1	7	-0.60	0.00
Intensity	3.23	1.76	131	1	7	0.34	-0.88
Reaction	2.32	1.68	131	1	7	1.15	0.34
Valence	-0.77	0.88	131	-3	1	-0.46	-0.06
See	5.69	1.15	131	2	7	-0.78	0.22
Setting	5.66	1.19	131	2	7	-0.82	0.31
Hearing	2.79	1.95	131	1	7	0.62	-0.94
In Words	4.32	1.81	131	1	7	-0.39	-0.87
Story	5.43	1.28	131	1	7	-0.78	0.47

Table 10

Experiment 1 Interim Retrieval MCQ Ratings Descriptive Statistics

MCQ Question	Mean	*SD*	*n*	Min	Max	Skew	Kurtosis
Reliving	5.28	1.38	39	1	7	-0.97	1.01
Intensity	3.51	1.73	39	1	7	0.01	-1.15
Reaction	2.33	1.54	39	1	6	1.04	-0.17
Valence	-0.62	1.02	39	-3	2	0.09	-0.02
See	5.97	0.96	39	3	7	-1.00	0.85
Setting	5.74	1.31	39	2	7	-1.31	1.32
Hearing	2.87	2.08	39	1	7	0.68	-1.02
In Words	4.44	1.87	39	1	7	-0.65	-0.70
Story	5.74	0.99	39	3	7	-0.43	-0.25

Table 11

Experiment 1 EFT Future Birthday MCQ Ratings Descriptive Statistics

MCQ Question	Mean	*SD*	*n*	Min	Max	Skew	Kurtosis
Reliving	4.58	1.54	45	1	7	-0.70	-0.45
Intensity	3.36	1.77	45	1	7	0.07	-1.27
Reaction	2.47	1.59	45	1	6	0.79	-0.55
Valence	1.22	1.58	45	-2	3	-0.60	-0.71
See	5.04	1.55	45	1	7	-0.86	0.35
Setting	5.38	1.25	45	2	7	-0.72	-0.01
Hearing	3.98	1.75	45	1	7	-0.07	-1.10
In Words	4.20	1.74	45	1	7	-0.43	-0.84
Story	4.89	1.56	45	1	7	-0.67	-0.01

Table 12

Experiment 1 EFT Police Interview MCQ Ratings Descriptive Statistics

MCQ Question	Mean	*SD*	*n*	Min	Max	Skew	Kurtosis
Reliving	5.00	1.46	45	1	7	0.26	0.22
Intensity	3.42	1.54	45	1	7	-0.72	0.23
Reaction	2.62	1.71	45	1	7	0.77	-0.35
Valence	-0.49	0.87	45	-2	2	0.68	1.08
See	5.11	1.19	45	1	7	-0.92	2.09
Setting	5.18	1.23	45	2	7	-0.33	-0.41
Hearing	4.29	1.67	45	1	7	-0.28	-0.92
In Words	4.42	1.63	45	1	7	-0.46	-0.68
Story	5.11	1.42	45	2	7	-0.47	-0.67

Appendix H

Initial Interview Questions

Version A

1. What shoes was the woman in pants with a stroller wearing?

2. What type of glasses was the older woman wearing?

3. What speed do you estimate the cyclist was moving at when he approached the red light? (Misleading)

4. When the woman was talking at the cross walk, do you think the woman saw the thief before he stole her bag?

5. What did the woman do with her crying baby when she saw the thief with her bag? (Misleading)

6. Where did the thief go after he put the stolen item in his own bag? (Misleading)

7. What color was the parked car the thief ran past after stealing the bag?

8. As the thief ran away, he bumped into a woman walking out of the park; what was her reaction? (Misleading)

Version B

1. What shoes was the women in dark blue jeans with a stroller wearing? (Misleading)

2. What type of glasses was the older woman reading a newspaper wearing? (Misleading)

3. What speed do you estimate the cyclist was moving at when he approached the traffic light?

4. When the woman was talking to the older woman, do you think the woman saw the thief before he stole her bag? (Misleading)

5. What did the woman do with her baby when she saw the thief with her bag?

6. Where did the thief go after he took the stolen items?

7. What color was the parked car two men were sitting in when the thief ran past after stealing the bag? (Misleading)

8. As the thief ran away a woman was walking out of the park; what was her reaction?

Appendix I

Experiment 2 Data Screening

I conducted data screening for three different models. All models were generalized multi-level models. Within the proposed models, there were two levels: the question level and the participant level. I conducted data screening at both levels of analysis. Below, I discuss how the data were deemed appropriate for analyses given each of our levels of interest. I further included a section on data analyses for the models that investigated differences in misinformation production across interim task conditions, and the models which explore the impact of past and future oriented detail production on later memory accuracy for the original event.

I evaluated the data at the question and participant level for the EFT and interim retrieval conditions separately. This separate analysis allows for the evaluation of the impact of past and future focused details during event construction (interim retrieval, EFT) on final memory test memory accuracy and misinformation production. The EFT and interim retrieval conditions were split given that the interim retrieval condition did not report focusing on any future oriented details during memory retrieval. Thus, examining the impact of future focused details on memory accuracy and misinformation production across both conditions would be inappropriate given that one of the two conditions would only have zeros for that predictor variable.

Models Examining differences between all conditions

Data screening by participant for Accuracy. I first analyzed accuracy on the final memory test at the participant level. When averaged across participants, the average accuracy as a proportion was .41 (SD = .20). The skew was less than |1| (skew = 0.01) and the kurtosis was less than |3| (kurtosis = -0.43). When examining the z-scores of participants accuracy, no participants had a z-score greater than |3| suggesting that there were no participants that were

outliers for final memory test accuracy. When looking at accuracy across interim task conditions, participants in the interim retrieval condition had an average accuracy of 0.44 (SD = .49), participants in the EFT condition had an average accuracy of 0.41 (SD = .49), and participants in the filler condition had an average accuracy of 0.37 (SD = .48).

Data screening by question for Accuracy. Second, I examined the final memory test accuracy at the question level. When averaged across all questions, the average accuracy as a proportion was 0.41 (SD = 0.18). The skew was slightly over |1| (skew = 1.10) and kurtosis was less than |3| (kurtosis = 0.26). When examining the z-scores associated with accuracy by question, none were greater than |3| suggesting there was no question that functioned as an outlier for their average accuracy. I then examined the average accuracy for neutral questions (*M* = 0.47, SD = 0.49) and misleading questions (*M* = 0.35, SD = 0.48).

Data screening by participant for Misinformation Production. I first analyzed misinformation production on the final memory test at the participant level. When averaged across participants, the average misinformation produced as a proportion was 0.36 (*SD* = 0.17). The skew was less than |1| (skew = 0.28) and the kurtosis was less than |3| (kurtosis = -0.25). When examining the z-scores of participants misinformation production, one participant had a z-score greater than |3|. This participant was retained during the analysis given that their misinformation production was not outside of the realm of possibility. When looking at accuracy across interim task conditions, participants in the interim retrieval condition had an average misinformation production of 0.35 (SD = .48), participants in the EFT condition had an average misinformation production of 0.37 (SD = .48), and participants in the filler condition had an average misinformation production of 0.37 (SD = .48).

Data screening by question for Misinformation Production. Second, I examined the final memory test misinformation production at the question level. When averaged across all questions, the average misinformation production as a proportion was 0.37 (SD = 0.19). The skew was less than |1| (skew = -0.53) and kurtosis was less than |3| (kurtosis = -0.79). When examining the z-scores of misinformation production by question, none were greater than |3| suggesting there was no question that functioned as an outlier for their average misinformation production. I then examined the average misinformation production for neutral questions (*M* = 0.25, SD = 0.44) and misleading questions (*M* = 0.48, SD = 0.50).

Models examining the impact of future and past oriented details in the EFT condition on final memory test accuracy and misinformation production

Data screening by participant for Accuracy. When averaging final police accuracy across participants in the EFT condition, the average accuracy as a proportion was 0.41 (SD = .19). The skew was less than |1| (skew = -0.06) and the kurtosis was less that |3| (-0.31). When examining the z-scores of participants accuracy, no participants had a z-score greater than |3| suggesting that there were no participants that were outliers for final memory test accuracy.

Data screening by participant for past and future focused details. I then evaluated the number of past and future focused details reported by participants in the EFT condition. On average, participants reported 40.55 (*SD* = 38.38) past focused details and 60.39 (*SD* = 49.32) focused details when imagining the police interview. When examining the z-scores of past focused items by participant, only one participant had a z-score greater than |3|. When looking at the z-scores of future focused details by participants, three participants had z-scores greater than |3|. The participants were retained for analysis given that the number of details that they generated was still plausible given the constraints of the task. Importantly, past and future

focused details were found to significantly correlate with one another (r = -.25; t (132) = -2.95, p = .003). That said, this correlation was weakly negative, and not a concern for multicollinearity given that it is less than |.9|. Thus, future and past focused details were determined to be appropriate continuous predictor variables to include in the same model.

Data screening by question for Accuracy. When averaging final memory test accuracy across questions, the average accuracy as a proportion was 0.41 (*SD* = .20). The skew was slightly greater than |1| (skew = 1.01) and the kurtosis was less than |3| (kurtosis = 0.12). When examining the z-scores of each question by accuracy, no question had a z-score greater than |3| suggesting that there were no questions that were outliers for final memory test accuracy. Neutral questions had an average accuracy of 0.46 (SD = .50), while misleading questions had an average accuracy of 0.41 (SD = .49).

Data screening by participant for Misinformation Production. When averaged across participants, misinformation produced on the final memory test as a proportion was 0.37 (SD = .17). The skew was less than |1| (skew = 0.34) and the kurtosis was less than |3| (kurtosis = -0.51). When examining the z-scores of participants misinformation production, no participants in the EFT condition had a z-score greater than |3| suggesting that there were no participants that were outliers for final memory test misinformation production.

Data screening by question for Misinformation Production. When averaged across questions, the average misinformation production produced during the final memory test as a proportion was 0.37 (SD = .19). The skew was less than |1| (skew = -0.53) and the kurtosis was less than |3| (kurtosis = -0.79). When examining the z-scores of each question by misinformation production, no question had a z-score greater than |3| suggesting that there were no questions that were outliers for final memory test misinformation. Neutral questions had an average

misinformation production of 0.28 (SD = .45), while misleading questions had an average misinformation production of 0.46 (SD = .50).

Models examining the impact of future and past oriented details in the retrieval condition on final memory test accuracy and misinformation production

Data screening by participant for accuracy. When averaged across participants, final memory test accuracy as a proportion was 0.44 (*SD* = .20). The skew was less than |1| (skew = -0.06) and the kurtosis was less than |3| (kurtosis = -0.45). When examining the z-scores of participants accuracy, no participants had a z-score greater than |3| suggesting that there were no participants that were outliers for final memory test accuracy.

Data screening by participant for past focused details. On average, participants reported 93.45 (SD = 41.42) past focused details when remembering the original event. When evaluating z-scores of past focused details produced by participant, one participant had a z-score greater than |3|. This participant was retained given that the number of past focused details they produced was still possible given the constraints of the task. Given that no participants in the interim retrieval condition produced future focused details when remembering the original event, I only analyzed past focused details during this data pre-screening.

Data screening by question for accuracy. When averaged across questions, final memory test accuracy as a proportion was 0.44 (SD = 0.18). The skew was less than |1| (skew = -0.57) and the kurtosis was less than |3| (kurtosis = -1.22). When examining the z-scores of question accuracy, no question had a z-score greater than |3| suggesting that there were no questions that were outliers for final memory test accuracy. Neutral questions had an average accuracy of 0.50 (SD = 0.50), while misleading questions had an average accuracy of 0.38 (SD = 0.49).

Data screening by participant for misinformation production. When averaged across participants, the final memory test misinformation production as a proportion was 0.35 (SD = .16). The skew was less than |1| (skew = 0.35) and the kurtosis was less than |3| (kurtosis = 0.59). When examining the z-scores of participants misinformation production, one participant had a z-score greater than |3|. This was the same participant who had a z-score greater than |3| when evaluating the whole data set. This participant was still retained in the analysis given that their final memory misinformation production was not outside the realm of possibility.

Data screening by question for misinformation production. When averaged across questions, final police misinformation production as a proportion was 0.44 (SD = .18). The skew was slightly greater than |1| (skew = 1.12) and the kurtosis was less than |3| (kurtosis = 0.22). When examining the z-scores of question misinformation production, no question had a z-score greater than |3| suggesting that there were no questions that were outliers for final memory test misinformation production. Neutral questions had an average misinformation production of 0.23 (SD = 0.42), while misleading questions had an average misinformation production of 0.47 (SD = 0.50).

Evaluation of MCQ data for t-test analyses

I planned two sets of t-tests for the following analysis. The first set would compare the response to each MCQ question between participants in the EFT condition and the interim retrieval condition when they were imagining the police interview and remembering the original event, respectively. Given that the filler condition never engaged in either task, they could not be included within this comparison. This comparison was conducted to evaluate how participants reported the experience of imagining the police interview compared to remembering the original event. The second set of t-tests compared participants' ratings of imagining the police interview

compared to imagining their next birthday. Importantly, in Experiment 1, there were no differences in phenomenological experience of imagining one's next birthday compared to the police interview other than in emotional valence of the event. Below, I screen the data to make sure they fit the assumptions needed for t-tests.

The data were found to meet the assumption of independence given that no participants were repeated through a sample. Furthermore, participants were randomly assigned to experimental conditions within the study. I statistically evaluated each data set for normality, and equal variance. Normality was evaluated using qqplots, skew, and kurtosis of each variable (Tables 13-15). In the interim retrieval condition, participants' ratings of the remembered setting and reliving of the original event were skewed and non-normal. In the EFT condition, participants' ratings of valence and setting for their imagined birthdays were skewed and non-normal. In the EFT condition, the distribution of all participants rating for the imagined police interview was approximately normal. To accommodate the non-normal distribution of a few variables, I used the non-parametric version of a t-test, The Mann-Whitney U test for these analyses. I evaluated the heterogeneity of variances for each pair of planned comparisons with Levene's Test. Several planned comparisons were found to not share equal variances; thus, I decided to conduct all analyses as Welch's t-tests, which do not assume equal variances between groups.

Table 13

Experiment 2 Interim Retrieval MCQ Ratings Descriptive Statistics

MCQ Question	Mean	*SD*	*n*	Min	Max	Skew	Kurtosis
Reliving	5.33	1.44	150	1	7	-1.00	0.80
Intensity	4.04	1.98	150	1	7	-0.08	-1.21
Reaction	2.83	1.87	150	1	7	0.70	-0.72
Valence	-0.70	1.38	150	-3	3	0.54	0.41
See	5.89	1.11	150	1	7	-1.48	3.15
Setting	5.89	0.97	150	3	7	-0.66	0.05
Hearing	3.10	2.18	150	1	7	0.50	-1.25
In Words	4.70	1.74	150	1	7	-0.53	-0.66
Story	5.81	1.11	150	1	7	-0.96	1.33

Table 14

Experiment 2 EFT Future Birthday MCQ Ratings Descriptive Statistics

MCQ Question	Mean	*SD*	*n*	Min	Max	Skew	Kurtosis
Reliving	5.46	1.67	134	2	7	-0.56	-1.23
Intensity	4.51	1.67	134	1	7	-0.46	-0.57
Reaction	3.14	1.85	134	1	7	0.39	-1.03
Valence	1.54	1.53	134	-3	3	-1.06	0.46
See	5.87	1.07	134	2	7	-0.74	0.09
Setting	5.98	1.24	134	1	7	-1.25	1.36
Hearing	5.14	1.62	134	1	7	-0.88	0.26
In Words	5.12	1.54	134	1	7	-0.79	0.23
Story	5.71	1.07	134	3	7	-0.53	-0.61

Table 15

Experiment 2 EFT Police Interview MCQ Ratings Descriptive Statistics

MCQ Questions	Mean	*SD*	*n*	Min	Max	Skew	Kurtosis
Reliving	5.43	1.34	134	1	7	-0.70	0.11
Intensity	4.36	1.76	134	1	7	-0.29	-0.67
Reaction	3.56	1.99	134	1	7	0.12	-1.25
Valence	-0.28	1.23	134	-3	3	0.23	0.36
See	5.63	1.16	134	2	7	-0.68	0.45
Setting	5.51	1.32	134	2	7	-0.76	0.01
Hearing	5.12	1.46	134	1	7	-0.86	0.68
In Words	4.87	1.59	134	1	7	-0.65	-0.27
Story	5.61	1.16	134	1	7	-0.70	0.62

Appendix J

Experiment 2 Final Memory Test Questions

1. What kind of pants was the woman with the baby wearing?

2. What was the older woman reading when she approached the woman with the baby?

3. What color was the traffic light facing the cyclist?

4. Where did the thief put the stolen items?

5. Who was the woman talking to when her purse was stolen?

6. What was the baby doing when the woman ran after the thief?

7. How many people were in the parked car that the thief ran past after stealing the bag?

8. Who ran into a woman leaving the park?

Appendix K

EFT Perceptual Focus Instructions

Since you are an eyewitness to the crime, you have been contacted by police to be interviewed about what you saw. Today, you will write about what you think will happen during the police interview tomorrow. Please imagine what giving your statement to police tomorrow will be like in as much detail as possible. There are TWO main components of imagining the police interview that are important for you as a witness. First, and similar to how you imagined your next birthday, you should imagine sensory details you expect to experience during the police interview (i.e., seeing, touching, tasting, hearing, smelling). Second, you should also imagine how you will answer questions about the crime and the sensory details from the crime you will report.

An example of imaging a conversation with someone and focusing on sensory details is as follows: The man will have a big brown beard. I will probably hear the ticking clock on the wall as he waits for me to answer him. The room will be sparsely decorated and smell like lemon floor cleaner. I will answer his question and say that I had been in a large grocery store yesterday. The store was so crowded, I was distracted by the colorful winter jackets and loud children running around.

Like in the above example, focus on imagined sensory details during the police interview and in your reflection on what happened during the crime. When imagining the police interview your imagination needs to include BOTH components listed above.

The 6-minute timer will start immediately when you advance to the next page. If you finish writing before the 6 minutes are up, please go back and continue to add as much information as possible, either relating to what you already wrote or what you may not have written about yet.

Please advance to the next page when you are ready to start imagining the police interview.

1. Based on the instructions you just read above, what is the **first** component for you to imagine about the police interview?

   Please re-read the instructions above if you are not sure.

   A. The sensory details you will experience while being interviewed by police (i.e., what you expect to hear, touch, taste, smell see)
   B. The emotional details you will experience while being interviewed by police (i.e., experiencing emotions, reflecting on the meaning of this event to you)
   C. Not Specified

2. Based on the instructions you just read above, what is the **second** component for you to imagine about the police interview?

   Please re-read the instructions above if you are not sure.

   A. The questions that police would ask about the crime and the sensory details you would respond with
   B. The details about what would happen after the police interview in the future
   C. Not Specified

3. Based on the instructions you just read above, what components should you include in your written imagination of the police interview?

   Please re-read the instructions above if you are not sure.

   A. The first component
   B. The second component
   C. Both components

4. Based on the instructions, what should you do if you finish writing before the 6 minutes are up?

   If you are unsure of the answer, please review the instructions above.

   A. Wait for the timer to be up

   B. Go back and continue to add as much information relating to the event as possible

   C. Not specified

----- PAGE BREAK -----

Points to focus on while imagining:
1. Sensory details about what you expect to happen during the police interview (i.e., what you expect to hear, touch, taste, smell, see)
2. The questions that police would ask about the crime and the sensory details you would respond with

Write your imagined police interview below.

If you finish writing before the timer is up, please go back and continue to add as much information as possible either relating to what you already wrote or what you may not have written about yet. *Please reference the list of details to focus on above if you are unsure what to write and include both components in your response.*

Appendix L

EFT Aperceptual Focus Instructions

Since you are an eyewitness to the crime, you have been contacted by police to be interviewed about what you saw. Today, you will write about what you think will happen during the police interview tomorrow. Please imagine what giving your statement to police tomorrow will be like in as much detail as possible. There are TWO main components of imagining the police interview that are important for you as a witness. First, and similar to how you imagined your next birthday, you should imagine emotional details you expect to experience during the police interview (i.e., experiencing emotions, reflecting on the meaning of this event to you). Second, you should also imagine how you will answer questions about the crime and the emotional details from the crime you will report.

An example of imaging a conversation with someone and focusing on emotional details is as follows: I will feel intimidate by the man in front of me. The ticking clock will stress me out as I think of how to respond to him. I will be fine in the room we are talking in. I will happily answer the man because I want to be helpful. I will say that I was feeling overwhelmed in the grocery store and didn't see much because when I am overwhelmed, I am easily distracted.

Like in the above example, focus on your imagined thoughts and emotions during the police interview and in your reflection on what happened during the crime. When imagining the police interview your imagination needs to include BOTH components listed above.

The 6-minute timer will start immediately when you advance to the next page. If you finish writing before the 6 minutes are up, please go back and continue to add as much information as possible, either relating to what you already wrote or what you may not have written about yet.

Please advance to the next page when you are ready to start imagining the police interview.

1. Based on the instructions you just read above, what is the **first** component for you to imagine about the police interview?

   Please re-read the instructions above if you are not sure.

   A. The sensory details you will experience while being interviewed by police (i.e., what you expect to hear, touch, taste, smell see)
   B. The emotional details you will experience while being interviewed by police (i.e., experiencing emotions, reflecting on the meaning of this event to you)
   C. Not Specified

2. Based on the instructions you just read above, what is the **second** component for you to imagine about the police interview?

   Please re-read the instructions above if you are not sure.

   A. The questions that police would ask about the crime and the emotional details you would respond with
   B. The details about what would happen after the police interview in the future
   C. Not Specified

3. Based on the instructions you just read above, what components should you include in your written imagination of the police interview?

   Please re-read the instructions above if you are not sure.

   A. The first component
   B. The second component
   C. Both components

4. Based on the instructions, what should you do if you finish writing before the 6 minutes are up?

   If you are unsure of the answer, please review the instructions above.

   A. Wait for the timer to be up

   B. Go back and continue to add as much information relating to the event as possible

   C. Not specified

----- PAGE BREAK -----

Points to focus on while imagining:

1. Emotional details about what you expect to happen during the police interview (i.e., experiencing emotions, reflecting on the meaning of this event to you)

2. The questions that police would ask about the crime and the emotion details you would respond with

Write your imagined police interview below.

If you finish writing before the timer is up, please go back and continue to add as much information as possible either relating to what you already wrote or what you may not have written about yet. *Please reference the list of details to focus on above if you are unsure what to write and include both components in your response.*

Appendix M

Experiment 3 Data Screening

The first model examined variation in memory accuracy on the final memory test while the second model examined variation in misinformation production on the final memory test. Both models included participant level and question level predictor variables. Thus, data screening was completed at the participant and question level for both models. The data was deemed suitable for both models and the results can be found below.

Final memory test Accuracy Data Screening

Data screening by question. First, I examined the final memory test accuracy at the question level. When averaged across all questions, the average accuracy as a proportion was 0.41 (*SD* = 0.18). The skew was less than |1| (skew = 0.94) and kurtosis was less than |3| (kurtosis -0.10). When examining the z-scores of accuracy by question, none were greater than |3| suggesting there was no question that functioned as an outlier for their average accuracy. I then examined the average accuracy for neutral questions (*M* = 0.49, *SD* = 0.50) and misleading questions (*M* = 0.37, *SD* = 0.47).

Data screening by participant. First, I examined the accuracy on the final memory test by participant (*M* = 0.41, *SD* = 0.19). The skew was less than |1| (skew = 0.10) and the kurtosis was less than |3| (kurtosis = -0.03) suggesting that the data was not problematically distributed for a generalized multi-level model at the participant level. Furthermore, an analysis of z-score of accuracy by participant demonstrated that there were no participants that were considered outliers (z-score > |3|) for final memory test memory accuracy. I then analyzed the data by condition. Participants in the perceptual condition (*M* = 0.42, *SD* = 0.49) had numerically similar memory accuracy compared to the aperceptual condition (*M* = 0.41, *SD* = 0.49).

Second, I analyzed the number of past and future oriented details produced during the interim task. On average, participants produced 30.85 past focused details (*SD* = 33.58) and 62.90 future focused details (*SD* = 43.48). When analyzing z-scores, there were two participants that had past oriented detail z-scores and future oriented detail z-scores that were greater than |3|. These participants were retained within the analysis, because their scores were not deemed incapable of generating within the time frame of the experiment.

Misinformation Production Data Screening

Data screening by question – Misinformation Production. First, I examined the misinformation production at the question level. When averaged across all questions, the average misinformation production as a proportion was 0.37 (*SD* = 0.17). The skew was less than |1| (skew = -0.21) and kurtosis was less than |3| (kurtosis = -1.55). When examining the z-scores of misinformation production by question, none were greater than |3| suggesting there was no question that functioned as an outlier for their average misinformation production. I then examined the average misinformation production for neutral questions (*M* = 0.24, *SD* = 0.43) and misleading questions (*M* = 0.50, *SD* = 0.50).

Data screening by participant – Misinformation Production. The average misinformation produced by participant as a proportion was 0.37 (*SD* = 0.17). When examining the skew and kurtosis of misinformation production by participant, the skew was less than |1| (skew = 0.41) and the kurtosis was less then |3| (kurtosis = 0.91). When examining the misinformation production across conditions, participants in the perceptual condition had a numerically similar proportion of misinformation production (*M* = 0.36, *SD* = 0.48) compared to participants in the aperceptual condition (*M* = 0.37, *SD* = 0.48). When evaluating the z-score of misinformation production by participant, there were three participants that had a z-score slightly above |3|. I

retained these participants within the model given that their misinformation production was possible to produce.

Data screening for ANOVAs

I assessed the MCQ data for their suitability for an ANOVA analysis. The data was confirmed to meet the assumption of independence and random sampling given that participants were randomly assigned to experimental conditions and participants only participated in the study once. To assess normality, I evaluated each MCQ question on a qqplot as well as their skew and kurtosis. All MCQ questions approximately followed the normal distributions of the data. Furthermore, none of the questions had a skew that was greater than |1| or a kurtosis greater than |3| (Kline, 2011; See Table 16). A bartletts test was not used to assess normality given that this test is extremely sensitive to larger sample sizes. I assessed homogeneity of variance for each planned ANOVA through Levene tests. Only two MCQ questions were found to have heterogeneity of their variances: Valence and Setting. ANOVAs can be robust to violations of the homogeneity of variance in large samples with the same sample size, but given our unequal sample sizes, I opted to conduct a non-parametric test, the Kruskal-Wallis test, for the analysis of valence and setting ratings within the MCQ.

Table 16

Experiment 3 MCQ Rating Descriptive Statistics

MCQ Questions	Mean	*SD*	*n*	Min	Max	Skew	Kurtosis
Reliving	5.36	1.32	271	1	7	-0.88	0.68
Intensity	4.60	1.68	271	1	7	-0.47	-0.60
Reaction	3.73	1.91	271	1	7	0.04	-1.13
Valence	-0.46	1.24	271	-3	3	0.50	0.04
See	5.40	1.30	271	1	7	-0.87	0.75
Setting	5.24	1.43	271	2	7	-0.73	-0.02
Hearing	4.83	1.58	271	1	7	-0.80	0.09
In Words	5.00	1.43	271	1	7	-0.72	0.20
Story	5.42	1.31	271	1	7	-0.88	0.77

References

Addis, D. R., Sacchetti, D. C., Ally, B. A., Budson, A. E., & Schacter, D. L. (2009). Episodic simulation of future events is impaired in mild Alzheimer's disease. *Neuropsychologia*, *47*(12), 2660–2671. https://doi.org/10.1016/j.neuropsychologia.2009.05.018

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-Related Changes in the Episodic Simulation of Future Events. *Psychological Science*, *19*(1), 33–41. https://doi.org/10.1111/j.1467-9280.2008.02043.x

Arnold, K. M., & McDermott, K. B. (2013). Free Recall Enhances Subsequent Learning. *Psychonomic Bulletin & Review*, *20*(3), 507–513. https://doi.org/10.3758/s13423-012-0370-3

Arnold, K. M., McDermott, K. B., & Szpunar, K. K. (2011). Imagining the near and far future: The role of location familiarity. *Memory & Cognition*, *39*(6), 954–967.

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

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a Witnessed Event Increases Eyewitness Suggestibility: The Reversed Testing Effect. *Psychological Science*, *20*(1), 66–73. https://doi.org/10.1111/j.1467-9280.2008.02245.x

D'Argembeau, A., Ortoleva, C., Jumentier, S., & Van der Linden, M. (2010). Component processes underlying future thinking. *Memory & Cognition*, *38*(6), 809–819. https://doi.org/10.3758/MC.38.6.809

D'Argembeau, A., Renaud, O., & Van der Linden, M. (2011). Frequency, characteristics and functions of future-oriented thoughts in daily life. *Applied Cognitive Psychology, 25*(1), 96–103. https://doi.org/10.1002/acp.1647

D'Argembeau, A., & Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness and Cognition, 13*(4), 844–858. https://doi.org/10.1016/j.concog.2004.07.007

D'Argembeau, A., & Van der Linden, M. (2006). Individual differences in the phenomenology of mental time travel: The effect of vivid visual imagery and emotion regulation strategies. *Consciousness and Cognition, 15*(2), 342–350. https://doi.org/10.1016/j.concog.2005.09.001

D'Argembeau, A., & Van der Linden, M. (2012). Predicting the phenomenology of episodic future thoughts. *Consciousness and Cognition, 21*(3), 1198–1206. https://doi.org/10.1016/j.concog.2012.05.004

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior, 33*(4), 298–307. https://doi.org/10.1007/s10979-008-9146-8

Gallo, D. A., Korthauer, L. E., McDonough, I. M., Teshale, S., & Johnson, E. L. (2011). Age-related positivity effects and autobiographical memory detail: Evidence from a past/future source memory task. *Memory, 19*(6), 641–652. https://doi.org/10.1080/09658211.2011.595723

Giner-Sorolla, R., Montoya, A. K., Reifman, A., Carpenter, T., Lewis, N. A., Aberson, C. L., Bostyn, D. H., Conrique, B. G., Ng, B. W., Schoemann, A. M., & Soderberg, C. (2024). Power to Detect What? Considerations for Planning and Evaluating Sample Size. *Personality and Social Psychology Review*, *28*(3), 276–301. https://doi.org/10.1177/10888683241228328

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

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

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

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Greene, C. M., Bradshaw, R., Huston, C., & Murphy, G. (2022). The medium and the message: Comparing the effectiveness of six methods of misinformation delivery in an eyewitness memory paradigm. *Journal of Experimental Psychology: Applied*.

Horselenberg, R., Merckelbach, H., Muris, P., Rassin, E., Sijsenaar, M., & Spaan, V. (2000). Imagining fictitious childhood events: The role of individual differences in imagination

inflation. *Clinical Psychology & Psychotherapy*, *7*(2), 128–137. https://doi.org/10.1002/(SICI)1099-0879(200005)7:2<128::AID-CPP238>3.0.CO;2-Q

Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: Evidence from semantic dementia. *Brain*, *135*(7), 2178–2191. https://doi.org/10.1093/brain/aws119

Irish, M., & Piolino, P. (2016). Impaired capacity for prospection in the dementias – Theoretical and clinical implications. *British Journal of Clinical Psychology*, *55*(1), 49–68. https://doi.org/10.1111/bjc.12090

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*(3), 306–340. https://doi.org/10.1037/0096-3445.110.3.306

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

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*(4), 371–376. https://doi.org/10.1037/0096-3445.117.4.371

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive Operations and Decision Bias in Reality Monitoring. *The American Journal of Psychology*, *94*(1), 37–64. https://doi.org/10.2307/1422342

Karpicke, J. D. (2017). Retrieval-Based Learning: A Decade of Progress. In *Learning and Memory: A Comprehensive Reference* (pp. 487–514). Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014a). Retrieval-Based Learning. In *Psychology of Learning and Motivation* (Vol. 61, pp. 237–284). Elsevier. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014b). Retrieval-Based Learning: An Episodic Context Account. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 61, pp. 237–284). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). Guilford Press.

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

LaPaglia, J. A., & Chan, J. C. K. (2013). Testing increases suggestibility for narrative-based misinformation but reduced suggestibility for question-based misinformaiton. *Behavioral Sciences and the Law*, *31*. https://doi.org/10.1002/bsl.2090

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

McCarty, R. (2016). Chapter 4 - The Fight-or-Flight Response: A Cornerstone of Stress Research. In G. Fink (Ed.), *Stress: Concepts, Cognition, Emotion, and Behavior* (pp. 33–37). Academic Press. https://doi.org/10.1016/B978-0-12-800951-2.00004-2

McDonough, I. M., & Gallo, D. A. (2010). Separating past and future autobiographical events in memory: Evidence for a reality monitoring asymmetry. *Memory & Cognition*, *38*(1), 3–12. https://doi.org/10.3758/MC.38.1.3

McGaugh, J. L. (2000). Memory—A Century of Consolidation. *Science*, *287*.

McLelland, V. C., Devitt, A. L., Schacter, D. L., & Addis, D. R. (2015). Making the future memorable: The phenomenology of remembered future events. *Memory*, *23*(8), 1255–1263. https://doi.org/10.1080/09658211.2014.972960

Memon, A., Wark, L., Bull, R., & Koehnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, *88*(2), 179–197. https://doi.org/10.1111/j.2044-8295.1997.tb02629.x

Newby-Clark, I. R., & Ross, M. (2003). Conceiving the Past and Future. *Personality and Social Psychology Bulletin*, *29*(7), 807–818. https://doi.org/10.1177/0146167203029007001

Paddock, J. R., Joseph, A. L., Chan, F. M., Terranova, S., Manning, C., & Loftus, E. F. (1998). When guided visualization procedures may backfire: Imagination inflation and predicting individual differences in suggestibility. *Applied Cognitive Psychology*, *12*(7), S63–S75. https://doi.org/10.1002/(SICI)1099-0720(199812)12:7<S63::AID-ACP600>3.0.CO;2-S

Painter, J. M., & Kring, A. M. (2015). Back to the Future: Similarities and Differences in Emotional Memories and Prospections. *Applied Cognitive Psychology*, *29*(2), 271–279. https://doi.org/10.1002/acp.3105

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*, *5*. https://doi.org/10.3389/fpsyg.2014.00286

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

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rubin, D. C. (2014). Schema-driven construction of future autobiographical traumatic events: The future is much more troubling than the past. *Journal of Experimental Psychology: General*, *143*(2), 612–630. https://doi.org/10.1037/a0032638

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 773–786. https://doi.org/10.1098/rstb.2007.2087

Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic future thinking: Mechanisms and functions. *Current Opinion in Behavioral Sciences*, *17*, 41–50. https://doi.org/10.1016/j.cobeha.2017.06.002

Schacter, D. L., Gaesser, B., & Addis, D. R. (2013). Remembering the Past and Imagining the Future in the Elderly. *Gerontology*, *59*(2), 143–151. https://doi.org/10.1159/000342198

Song, H., & Schwarz, N. (2008). Fluency and the Detection of Misleading Questions: Low Processing Fluency Attenuates the Moses Illusion. *Social Cognition*, *26*(6), 791–799. https://doi.org/10.1521/soco.2008.26.6.791

Suengas, A. G., & Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General*, *117*(4), 377–389. https://doi.org/10.1037/0096-3445.117.4.377

Szpunar, K. (2010). Episodic Future Thought: An Emerging Concept. *Perspectives on Psychological Science*, *5*(2), 142–162. https://doi.org/10.1177/1745691610362350

Szpunar, K. K., & McDermott, K. B. (2008). Episodic future thought and its relation to remembering: Evidence from ratings of subjective experience. *Consciousness and Cognition*, *17*(1), 330–334. https://doi.org/10.1016/j.concog.2007.04.006

Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language*, *63*(2), 149–157. https://doi.org/10.1016/j.jml.2010.04.004

Thomas, A. K., Bulevich, J. B., & Loftus, E. F. (2003). Exploring the role of repetition and sensory elaboration in the imagination inflation effect. *Memory & Cognition*, *4*(31), 630–640.

Thomas, A. K., Gordon, L. T., Cernasov, P. M., & Bulevich, J. B. (2017). The effect of testing can increase or decrease misinformation susceptibility depending on the retention interval. *Cognitive Research: Principles and Implications*, *2*(1), 45. https://doi.org/10.1186/s41235-017-0081-4

Tomes, J. L., & Katz, A. N. (1997). Habitual Susceptibility to Misinformation and Individual Differences in Eyewitness Memory. *Applied Cognitive Psychology*, *11*(3), 233–251. https://doi.org/10.1002/(SICI)1099-0720(199706)11:3<233::AID-ACP447>3.0.CO;2-V

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

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, *18*(6), 1140–1147. https://doi.org/10.3758/s13423-011-0140-7