Does Warning Participants Attenuate the Retrieval-Enhanced Suggestibility Effect? Exploring a Retrieval Fluency Explanation

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Psychology

at Massey University, Manawatū,

New Zealand.

Amanda Naus

`

2014

Abstract

The purpose of the current research was to explore a retrieval fluency explanation for the retrieval-enhanced suggestibility (RES) effect and to investigate whether the RES effect can be reduced with a warning. Participants were a volunteer sample of 91 people between 18 and 45 from across New Zealand, recruited via social media and word of mouth. A 3 x 2 x 2 factorial design was used, with participants being assigned to either the single testing or repeated testing condition, and either the warning or no warning condition. Item type (misinformation, consistent, or control) was manipulated within-subjects. Participants watched a video, then either took a 15 question forced-choice test or played the video game Tetris. All participants then read a summary of the video event, which contained some misleading information. Participants in the warning conditions were then warned the information may not be accurate. After watching a distractor video, all participants completed the same 15 question forced-choice test that has been used earlier. Average hit rates, confidence ratings, and response latencies were measured. It was hypothesised that participants in the repeated testing condition would report more of the misinformation and consistent items on the final test, compared to participants in the single testing condition, but only for participants who were not warned. It was also hypothesised that participants in the repeated testing condition would have higher confidence, and faster response latencies, for misinformation and consistent items, than participants in the single testing condition, but only for participants who were not warned. Results showed there was limited evidence for the RES effect, and some evidence for a response

fluency-based explanation for the RES effect. The mixed support for the various hypotheses was discussed along with the limitations of the current research, and recommendations for future research.

Acknowledgements

I am using this opportunity to express my sincere gratitude to everyone who supported me throughout this research project. Firstly, I would like to thank my supervisor, Dr. Stephen Hill. Without his assistance and dedicated involvement throughout every step of the process, this project would not have been possible. Thank you for your guidance, encouragement and expert advice.

I would also like to thank Malcolm Loudon for his expertise in developing the technical applications necessary for this project to proceed.

I would like to thank my husband, Phillip, for his endless support and encouragement. Finally, I would like to thank all of the family and friends who supported me, and who were kind enough to lend their time to take part in this study, thereby contributing to this important field of psychology.

The materials and procedures used in this project were evaluated by peer review and judged to be low risk. Consequently, this project was not reviewed by the Massey University Human Ethics Committee.

Table of Contents

List of Tables

CHAPTER ONE

Introduction

In criminal trials, the accuracy of the eyewitness account is often the most important factor in whether the case will be solved (Cutler & Penrod, 1995). However, decades of research have shown that eyewitness memory is highly susceptible to errors and interference from misleading information (see, eg., Loftus, 2005). In fact, inaccurate eyewitness testimonies are the most common reason for the miscarriage of justice in criminal cases (Gieselman, Fisher, MacKinnon, & Holland, 1986; Loh, 1981). When a person witnesses an event, they are almost certain to receive later information relating to that event. This information may be inaccurate, and inaccurate information can causes the person's memory for that event to become distorted. It is well established that receiving misinformation after an event can distort the memory for that event. This phenomenon is widely known as the misinformation effect (Loftus, 2005). Loftus (2005) provided a comprehensive review of the findings of research about the effect of misinformation on eyewitness memory over the last 30 years. However, the majority of research on the misinformation effect does not take into account the fact that in real-life situations, eyewitnesses often recall the event several times prior to exposure to misinformation. For example, after witnessing a crime, an eyewitness will usually talk about the incident with fellow witnesses, the media, or with responding emergency personnel, thus recalling the event (Chan, Thomas, & Bulevich, 2009). Paterson and Kemp (2006) report that eyewitnesses are likely to talk to each other following an event, before they are interviewed by the police. Their memory for the event is therefore tested, often before they are exposed to misinformation. The misinformation they are later exposed to may come from the fellow witnesses, or from more trustworthy sources such as the media coverage of the event (Allen & Lindsay, 1998; Lindsay, Allen, Chan, & Dahl, 2004).

It is also well established that if memory is tested after an event, the memory for that event is strengthened (Roediger & Karpicke, 2006). However, Chan et al. (2009) found that when participants were tested prior to being exposed to misleading information, they were more likely to recall the misinformation on a later test. Thus, it seems that the testing of participants' memory for the event immediately afterwards somehow increased their assimilation of the misleading information. They termed this the retrieval-enhanced suggestibility effect (RES). The RES effect has been found in both younger and older adults (Chan et al., 2009), with both cued-recall and recognition (Thomas, Bulevich, & Chan, 2010), and the RES effect is enhanced when participants take multiple tests (Chan & LaPaglia, 2011). Thus, it is clearly a cause for concern with respect to the practical implications for the criminal justice system.

Thomas et al. (2010) investigated whether warning participants about the potential inaccuracy of the post-event information would decrease the RES effect. Previous research has shown results are mixed when it comes to the effectiveness of post-warnings on reducing eyewitness suggestibility (a detailed discussion follows in the next chapter), however, Thomas et al. (2010) found warning participants eliminated the RES effect, thus providing a strategy for overcoming the effect. To explain the RES effect and its reduction with a warning, Thomas et al. (2010) provide a retrieval fluency explanation. They argue that participants respond based on retrieval fluency, but when warned their basis of responding changes to a more effortful recollection process. This explanation will be discussed further in the following chapter.

The present study investigates whether the RES effect can be eliminated by warning participants about the potential inaccuracy of the information they have been presented with. The standard RES paradigm will be used, and retrieval fluency of participants' responses, along with their confidence levels, will be used to ascertain whether the retrieval fluency hypothesis holds. The following chapter will comprehensively review the literature in the area of the current research.

CHAPTER TWO

Literature Review

Introduction

Eyewitness accounts of criminal events are often the most important aspect of a case. While eyewitness memory has long been shown to be malleable, recent research by Chan et al. (2009) highlighted an interesting phenomena: that when participants are tested prior to exposure to misinformation, they are then more susceptible to that misinformation on a final test. Because previous research has shown that testing leads to increased retention of information (see Roediger & Karpicke, 2006), this finding was somewhat surprising. The current literature review focuses on several areas of research related to the RES effect. The first part reviews the main findings from the large body of research looking at the how misleading information can become assimilated into a person's memory. We look at what the general misinformation paradigm is, what makes it more likely to occur, and who may be more susceptible to the misinformation effect than others. The second part provides a brief review on the vast literature on the testing effect. Following this is a review of the RES literature, including the main findings from the research so far, and what mechanisms have been proposed to cause the RES effect. The literature on warnings of possible inaccuracies in information provided, a possible strategy for reducing the RES effect, is then reviewed, with respect to what types of warning there are and under what conditions warnings may reduce the misinformation effect. The literature review then discusses the research relating to the retrieval fluency hypothesis, which is directly relevant to the current research.

Misinformation Effect

The misinformation effect has been well researched over many decades (Loftus, 2005, provides a comprehensive review of misinformation research, see also Ayers & Reder, 1998; Zaragoza, Belli, & Payment, 2007). The misinformation effect occurs when exposure to misleading information after learning results in an impairment of memory for the initial material (Loftus, 2005). For example, an eyewitness to a vehicle accident may falsely remember the presence of another car after this has been suggested by another witness. The numerous experiments that have been carried out in an attempt to better understand the misinformation effect have largely utilised the misinformation paradigm, developed by Loftus, Miller, and Burns (1978). In such experiments, participants generally witness some sort of event, either a simulation or a video, and are then subject to some misleading post-event information. This information either supplements or contradicts details of the original event. They then undergo a memory test where their memory for the original event is tested. Where the misinformation effect comes in to play is when participants answer the questions with details from the post-event information rather than the details of the original event.

Misinformation in real-life situations can come from many sources. It has been shown that when co-witnesses to an event discuss details of the event together, incorrect information suggested by one witness can become incorporated into the other eyewitness's memory for that event. This phenomenon is known as memory conformity (Gabbert, Memon, & Allen, 2003). Whether or not misinformation intrudes into a person's memory for the event depends on several factors, including the time delay between the event and the introduction of misinformation, and the age of the eyewitness. These factors will be discussed below.

The misinformation effect is increasingly likely to occur the more time has passed between the event and the introduction of misinformation (Loftus et al., 1978). One explanation for this is that the more time has passed, the less likely one is to notice that the misinformation does not match the details they remember about the original event. Thus, the person receiving the misinformation accepts it as part of their original memory. However, even if a discrepancy between the misinformation and the original event is noticed, the person may take that to mean that their memory for the original event is incorrect, thus they may accept the misinformation regardless of the discrepancy (eg., Loftus & Hoffman, 1989). Another factor that increases the likelihood of the misinformation effect is the time interval between the misinformation and the test. Higham (1998) asked participants whether a key detail was in the original event, the misinformation presented later, in both, or in neither. It was found that with a short interval between the misinformation and the test, participants were less likely to report the key detail as being part of the original event only. Therefore, with a short interval they are more likely to remember seeing the key detail as part of the misinformation, but they may also believe the detail was part of the original event.

It is evident that regardless of the time between the event and the intrusion of misinformation, some people are more susceptible to the misinformation effect than others. Age is one factor that influences how one experiences the misinformation effect. For instance, young children are more susceptible than older children and adults (see, for example, Ceci & Bruck, 1993). Additionally, older adults are more susceptible than younger adults (Karpel, Hoyer, & Toglia, 2001; David & Loftus, 2005; Roediger & Geraci, 2007).

Testing effect

The testing effect describes the phenomenon where being tested after learning new material results in increased accuracy of retrieval on a final test (see, eg., Roediger & Karpicke, 2006). Generally, the materials used in testing effect research involve word lists (e.g., Hogan & Kintsch, 1971; Izawa, 1967; McDaniel & Masson, 1985; Thompson, Wenger, & Bartling, 1978; Tulving, 1967; Wheeler, Ewers, & Buonanno, 2003) or picture lists (Wheeler & Roediger, 1992). However, Roediger and Karpicke (2006) used educationally-relevant materials (prose materials and free-recall tests without feedback) and looked at how the benefit of repeated testing compared to the benefit of restudying the material. Participants read a short passage and then either took a free-recall test or studied the passage again. It was found that while studying the passage improved recall, in comparison to initial testing, after a 5 minute delay, the reverse effect was found on delayed tests (Roediger & Karpicke, 2006). That is, while studying material yields a short-term benefit in terms of retention, testing improves retention in the longer-term. In a second experiment, participants either read the passage and then took three tests, or read the passage and studied it three more times. Those in the repeated-testing condition recalled far more than those in the repeated-study condition (61% vs. 40%) after a week–long delay, even though those in the repeated-study condition had had increased exposure to the material.

Thus, testing has a clear benefit to long-term retention, and this benefit goes beyond simply having an opportunity to restudy the material. One plausible explanation for the testing effect is the transfer-appropriate processing theory. This theory states that the compatibility between the operations used in the learning and testing phases improves learning (Morris, Bransford, & Franks, 1977; Roediger, 1990). Essentially, when participants are tested in the learning phase, they are practicing the exact skill (recalling a prose passage) that is needed when it comes to the retrieval phase, so this leads to improved retention (Roediger, Gallo, & Geraci, 2002).

Retrieval-Enhanced Suggestibility Effect

As outlined above, the misinformation effect refers to an impairment in memory when material learned has been tainted by misleading information (Loftus, 2005). The testing effect is the increased retention of material on a final test, after memory has been tested (Roediger & Karpicke, 2006).The research on the testing effect (discussed above) suggests testing participants before they are exposed to misleading information might protect them from incorporating the misinformation into their memories. Based on what we know from the vast amount of research in both the misinformation literature and the testing effect literature, it would be expected that if two groups of people are exposed to the same misleading information, the group whose memory was tested prior to receiving the misleading information would have a more accurate recall of the initial material. However, from 2009, Chan and colleagues have been investigating a phenomenon that was the opposite of what was expected based on existing theory; that initial testing before being presented with misinformation actually makes participants *more* susceptible to that misinformation. The basic experimental paradigm is that participants watch a video event, and then either have their memory for that event tested, or carry out an unrelated task. All participants then listen to an audio narrative that purportedly summarises the video event, when in fact it contains several misleading facts about the event. Finally, all participants have their memory tested, using the same test as those in the test condition completed originally. Chan et al. (2009) found that, surprisingly, participants who took a repeated test recalled more of the misinformation items than participants who took a single test.

Chan et al. (2009) used a cued-recall test to examine the RES effect in younger adults. They found that repeated testing increased recall of misinformation, compared to single testing. They then carried out the same experiment with older adults. Older adults were found to show the same retrieval-enhanced suggestibility effect as the younger adults. Unlike the younger adults, however, they did show a testing effect in that the participants who took a repeated test showed an improvement in accurate recall of both control and consistent items in the final test, compared to participants who took a single test (Chan et al., 2009). That is, testing improved their recall for details in the video that were consistent with correct post-event information, as well as correct details in the video that were not mentioned in the post-event information, but for the details in the video that were contradicted in the post-event information (misinformation items) they exhibited the misinformation effect, recalling the misinformation lure rather than the correct detail.

While research in the area of RES is in its infancy, the RES effect has been found in both young and older adults (Chan et al., 2009, see above) and with both cued-recall and recognition (Thomas et al., 2010). Thus, while the area requires further research in order to determine what other situations the RES effect can be found in, and how it can be reduced, the RES effect is clearly a robust finding, and this has important implications for the criminal justice system.

Chan and LaPaglia (2011) found that taking multiple tests increased the RES effect. The average RES effect size when taking a single initial test was .30, as measured by Cohen's *d,* compared to an average effect size of .76 when taking five or six initial tests. Eyewitnesses are likely to be questioned about the event multiple times by police, as well as by friends and family, before they are called to give evidence in court. Thus, their memory for the event is being tested multiple times, and this suggests that they will be particularly susceptible to misinformation. Undergoing multiple tests before being subject to misinformation is more akin to real-life situations than undergoing only one, and if the RES effect is stronger after multiple tests then it is clear that relying on eyewitness memory in criminal justice cases in problematic. How can we prevent eyewitnesses to a crime from being increasingly susceptible to misinformation?

LaPaglia, Wilford, Rivard, Chan, and Fisher (2013) investigated whether the RES effect would be found under ideal questioning conditions. It was thought that perhaps if eyewitnesses were questioned under conditions similar to those prescribed by the National Institute of Justice's technical working group (Technical Working Group for Eyewitness Evidence, 1999), this may increase accurate recall and reduce susceptibility to misinformation from outside sources. The Cognitive Interview (CI) is a method of questioning which focuses on establishing rapport with the eyewitness, avoiding interruptions and distractions from the interviewer, and encouraging free recall while discouraging guessing (LaPaglia et al., 2013). The CI process involves encouraging the witness to recall the event many times and uses various forms of retrieval (ie., verbal recall, drawing a sketch). It has been shown that the CI is more effective at increasing the recollection of relevant details of a witnessed event without a subsequent loss of accuracy, compared to other commonly used methods including the standard police interview (e.g., Fisher, Geiselman, & Amador, 1989), the structured interview technique (e.g., Akehurst, Milne, & Kohnken, 2003), and free recall (e.g., Gabbert, Hope, & Fisher, 2009). The CI can also reduce false recall. Geiselman, Fisher, Cohen, Holland, and Surtes (1986) found that participants who were exposed to misinformation recalled less of the misinformation following a CI, compared to a standard interview.

Following from these findings, LaPaglia et al. (2013) formed two alternative hypotheses. One hypothesis was that participants who completed the CI would be less susceptible to misinformation than participants who completed a free recall test or a distractor task, given that the CI has been shown to increase accurate recall, and uses techniques that have not been examined in previous RES research. A conflicting hypothesis is that participants who completed the CI would be more susceptible to misinformation than participants who completed a free recall test or a distractor task, therefore showing the RES effect. As taking multiple initial tests has been shown to produce a stronger RES effect than only taking one initial test, and the CI effectively involves multiple tests in that participants recall the event several times using different formats, it was thought that the CI might produce a strong RES effect. It was found that participants who completed the CI recall more of the misinformation in comparison to participants who completed a distractor task. Clearly, even the use of the ideal questioning conditions, as prescribed by the National Institute of Justice's technical working group (Technical Working Group for Eyewitness Evidence, 1999), does not protect eyewitnesses from the RES effect, and it fact leads to an increase in recall for misinformation. So what, if anything, can be done to eliminate, or at least reduce, the RES effect?

It has been discovered that the RES effect can be eliminated with a warning. Thomas et al. (2010) observed the RES effect when participants were not warned that the accuracy of the post-event information they were given could not be confirmed, but when participants were warned, the RES effect was eliminated. They argue that when participants take an initial test and then listen to the narrative containing misleading information, the details in the narrative that correspond to details present in the initial test (eg., a question in the test asks what colour the shirt one of the main characters was wearing in the video event, and the narrative contains misleading information about the colour of the shirt) capture their attention, which results in a deeper encoding of these details (Thomas et al., 2010). Because they are better encoded, the details therefore come to mind more quickly on the later test. Thus, this explains why people who take an initial test are more susceptible to the misinformation. Thomas et al. (2010) refer to this as the retrieval fluency hypothesis. They argue that participants' answers on the final test are based on the ease of retrieval of the information, and participants do not analyse the source of that information. They further argue that if participants are warned about the potential inaccuracy of the information, they will engage in a more effortful recollection process rather than responding based on ease of retrieval, and therefore the RES effect will be attenuated.

In order to test this hypothesis, they examined three things; whether warning participants about the potential inaccuracy of the post-event information would eliminate the RES effect, how confident participants were in their answers, and whether taking an initial test effected their retrieval latencies on the final test (Thomas et al., 2010). Because confidence is negatively related to retrieval latency (Kelley & Lindsay, 1993; Koriat, 1993; Koriat, Ma'ayan, & Nussinson, 2006), Thomas et al. (2010) reasoned that both confidence and response latency measures provided evidence for the retrieval fluency hypothesis in the context of RES. They argued that because initial testing enhances encoding of details and hence leads to easier retrieval later on, the misinformation and consistent items presented in the post-event information would come to mind more quickly on the final test when participants had taken an initial test. Because faster retrieval fluency is related to increased confidence in participants' answers (see discussion later on in this chapter), it was also thought that participants would have more confidence in their answers when initially tested, compared to participants who were not initially tested. Their results supported this hypothesis. Thomas et al. (2010) also measured response latency as a more direct measure of retrieval fluency, and found participants in the repeated testing condition responded faster than those in the single testing condition. Thus, evidence from confidence ratings and response fluency measures provides evidence for the retrieval fluency hypothesis, as it shows that participants respond based on how quickly the answer comes to mind. Although Thomas et al. (2010) found that giving a warning was effective in reducing susceptibility to misinformation, previous research on warnings has resulted in mixed findings. A review on the literature around whether a warning can reduce the misinformation effect is provided later on in this chapter.

Theoretical Explanations and Cognitive Mechanisms Underlying RES

There have been several theories used to explain exactly how the RES effect occurs. Chan et al. (2009) proposed two possible mechanisms that explain the RES effect. The first was the insulation effect, an idea that originated in the field of verbal learning (Tulving & Watkins, 1974). The insulation effect is the finding that when learning two sets of paired associates, recall for the second set is better if recall for the first has been tested before the second set is learned. Essentially, then, according to the insulation effect, learning of new information is enhanced by taking a test of previously learnt information. This could explain why participants who take a repeated test are more susceptible to later misinformation than those who took a single test, as found by Chan et al. (2009), as taking the initial test might have enhanced their learning of the misleading information that followed.

An alternative explanation put forward by Chan et al. (2009) is reconsolidation. Previous studies have found that reactivation of consolidated memories can return them to a labile state, where they must be reconsolidated, and at this time are more susceptible to interference (Nader, Schafe, & Le Doux, 2000; Walker, Brakefield, Hobson, & Stickgold, 2003). In terms of the RES effect, participants who were initially tested may have had their memory for the event reactivated because the initial test caused them to retrieve their memory for the original event, and thus were more susceptible to the misinformation in the post-event information. Memory retrieval has been shown to enhance assimilation of newly learned information with the memory for the original event (eg., Brown, Brown, Mosbacher, & Dryden, 2006; Henkel & Franklin, 1998; Hupbach, Gomez, Hardt, & Nadel, 2007; Zaragoza & Lane, 1994). Thus, the effect of the initial test might be to make the memory more susceptive to incorporating misinformation than would otherwise be the case.

To ascertain which of the above explanations could account for the RES effect, Chan et al. (2009) carried out a second experiment using the modified-modified free-recall (MMFR) design. The MMFR design was developed by Barnes and Underwood (1959) as a way to avoid response competition in situations where participants had to remember more than one piece of information associated with the same cue. In MMFR, participants are asked to recall everything they can that is associated with the cue. For example, when given a cue, participants in the Chan et al. (2009, Experiment 2) study could give all the information they could think of that is associated with that cue. If the cue related to a detail presented in a misleading way in the post-event information, participants could give both the original details and the misinformation detail. Thus, using MMFR, Chan and colleagues (2009) were able to investigate whether RES could be explained by the insulation effect or reconsolidation. If the insulation effect accounted for RES, then recall of misinformation would be increased after being tested, while recall for the initial event would be unaffected. Alternatively, if reconsolidation accounted for RES, the memory for the initial event would return to its labile state and be more susceptible to interference from the misinformation, therefore the memory for the original event would be less accurate and the memory for misinformation would be unaffected. Results showed that initial testing led to an improved recall of misinformation, but there was no difference in the accurate recall of misinformation items among participants in the single testing and the repeated testing conditions. Prima facie, results did not support the reconsolidation account, but this was thought to be due to a confound between the benefits of testing and the impact that interference had on the reconsolidation process. That is, participants in the repeated testing condition were negatively impacted by interference, but there was also an increase in accurate recall for misinformation items, which effectively clouded the interpretation of the reconsolidation account.

Chan and Langley (2011) attempted to elucidate the mechanisms underlying RES further, by manipulating the delay between the original event, the presentation of misinformation, and the final test. They carried out two experiments in order to find out whether the RES effect could be explained by the insulation effect, the reconsolidation effect, or both. The two experiments followed the standard RES procedure, with a seven-day retention interval introduced. In the first experiment (Experiment 1), participants watched the video event, completed an initial test or distractor task, listened to the information narrative, and then after a seven-day delay completed the final test. In the second experiment (Experiment 2), participants watched the video event, completed an initial test or distractor task and then after a seven-day delay listened to the information narrative, and completed the final test. These experiments were specifically designed to shed light on the mechanisms behind the RES effect. Specifically, if the reconsolidation account explains the RES effect (interference due to the original memory being reactivated by the initial test), then the RES effect would occur in Experiment 1 but not in Experiment 2. This is because a reactivated memory is only susceptible to interference for a short period of time, before the memory is reconsolidated (Judge & Quartermain, 1982). Thus, in Experiment 2, where the misinformation is presented seven days after the initial test, the memory should already be reconsolidated and thus the misinformation will not be assimilated into memory (Berman & Dudai, 2001; Przbyslawski, Roullet, & Sara, 1999). If the reconsolidation account is the mechanism behind the RES effect, then there should be no RES effect in Experiment 2. However, if the insulation effect is behind the RES effect (learning of new information is enhanced by taking a test of previously learnt material), there are two possible mechanisms that may explain this effect. Firstly, being tested on previously learnt material may reduce susceptibility to proactive interference (PI, defined as the tendency for existing memories to hinder the ability to retrieve memories learned later) when learning the misinformation (Szpunar, McDermott, & Roediger, 2008). If this is the case, then taking an initial test would enhance learning of the misinformation, thus leading to RES. Secondly, taking an initial test of several critical details of the original event may draw participants' attention to those details during the post-event information, which may cause them to pay more attention to those critical details during new learning (see, eg., Nelson, Dunlosky, Graf, & Narens, 1994; Thompson et al., 1978). Thus, taking a test may enhance participants' learning for the critical details in the post-event information, including misinformation, leading to an RES effect.

Chan and Langley (2011) designed the experiments so that they would also be able to distinguish which of these two mechanisms (reduced susceptibility to PI or an increase in attention to details highlighted in the test) would account for the enhanced learning of new material. If the initial test reduced susceptibility to PI, then the RES effect would only occur in Experiment 1. As the effect of PI on learning of new material decreases with delay (Underwood & Freund, 1968), then in Experiment 2, where there is a seven-day delay before misinformation is presented, it is likely that the initial test would no longer protect participants from PI, thereby enhancing learning of misinformation, as it would in Experiment 1. Alternatively, if RES is due to participants paying more attention to critical details in the post-event information due to their attention being drawn to those details that appeared in the test, then the RES would occur in both Experiment 1 and Experiment 2.

Chan and Langley (2011) found that the RES effect occurred in both Experiment 1 and Experiment 2, thus eliminating both the reconsolidation explanation and the test-reduced PI explanation. Thus, Chan and Langley (2011) argue that the RES occurs because the initial test asks questions about specific details of the witnessed event, and participants' attention is then drawn to those details, even if they are misleading, when they listen to the post-event information. This theory is consistent with the findings of Chan and LaPaglia (2011), who found that testing participants multiple times increased the RES effect significantly. Given that testing draws attention to particular details of the witnessed event, and causes participants to preferentially encode those details in any new information they are given, when taking multiple tests the details in the post-event information that relate to the tested details become even more memorable, leading to a larger RES effect.

Warning Participants about Possible Inaccuracies

As outlined above, Thomas et al. (2010) discovered that warning participants that the post-event information they received may be inaccurate significantly reduced the RES effect. The fact that giving participants a warning about the potential inaccuracy of information can eliminate the RES effect has two important implications. Firstly, there are important legal implications. If warning eyewitnesses that some of the information they have received may be inaccurate reduces the amount of misinformation they ‘remember’, then they will be more accurate eyewitnesses. Of course, this is an important objective for the justice system, and researchers have been investigating for years in an effort to meet this objective (for reviews, see eg., Wells & Olson, 2003). Secondly, there are theoretical implications. If warning participants causes them to remember the correct information rather than the misinformation, this implies that people must still have access to their memory for the original event after being exposed to misinformation. Thus, the warning must cause participants to consider the source of their memory, and therefore distinguish between the original event and any information received afterwards (see Christiaansen & Ochalek, 1983; Wright, 1993).

Such a finding is inconsistent with the theory put forward by Loftus and Hoffman (1989), that the misinformation essentially replaces the detail of the original memory, so that the misinformation becomes indistinguishable from memory of the original event. The effectiveness of warnings in reducing susceptibility to misinformation, and the associated theoretical implications, depends on whether participants are warned before they receive the misinformation, or after.

It is important to distinguish between pre-warnings and post-warnings. With a pre-warning, participants are warned before they receive the misinformation (Chambers & Zaragoza, 2001). Green, Flynn, and Loftus (1982) warned participants before they were exposed to misleading information that the information may be inaccurate, and found that this warning was effective in reducing suggestibility. While a pre-warning would be ideal, as participants could essentially protect themselves from assimilating the misinformation into their existing memory of the event by avoiding deep encoding of the information (Craik & Lockhart, 1972), it is often not possible in real-life situations as it would be difficult to warn eyewitnesses about potential misinformation before they receive it (Echterhoff, Hirst, & Hussy, 2005). This is because, after a crime, those responsible for investigating the event are unlikely to have access to the witnesses before they are exposed to post-event information.

With a post-warning, participants are warned after they receive the misinformation (Chambers & Zaragoza, 2001). The evidence for the effectiveness of post-warnings in reducing the effect of misinformation is mixed. Several studies have found the effect of misinformation to be fairly resistant to a post-warning (Belli, Lindsay, Gales, & McCarthy, 1994; Frost, Ingraham, & Wilson, 2002; Greene et al., 1982; Higham, 1998; Zaragoza & Lane, 1994, Experiment 4). Other studies have found that giving a warning can reduce the effect of misinformation (Chambers & Zaragoza, 2001; Highhouse & Bottrill, 1995; Meade & Roediger, 2002). Still further studies have found a warning can eliminate the effect of misinformation altogether (Blank, 1998; Wright, 1993; Zaragoza & Koshmider, 1989). In a meta-analysis of 25 post-warning studies, Blank and Launay (2014) found post-warnings reduced the misinformation effect by about half on average. The effectiveness of a post-warning in reducing the misinformation effect depends on several factors, including the specificity of the warning, the grammatical mood the warning is framed in, and the credibility of the person providing the warning.

While there are factors that may determine whether or not a warning is successful in reducing the effect of misinformation, it is difficult to compare many of the aforementioned studies as they differ somewhat in the type of misinformation (supplementary or contradictory), the type of memory test (source memory, recognition, or recall) and even in the type of post-warning (explicit or implicit). Warnings that explicitly point out that there are discrepancies between the event details and the post-event information (Wright, 1993) or frame the warning in an indicative mood (ie., stated as an actuality or fact; eg., “you saw the correct answer in the slide show, but an incorrect answer was mentioned in the story, Lindsay, 1990, p. 1080), are more effective than non-specific post-warnings (eg., Belli et al., 1994; Christiaansen & Ochalek, 1983; Frost et al., 2002; Zaragoza & Lane, 1994), or warnings framed in a subjunctive mood (ie., stated as a doubtful condition; eg., “there may have been some crucial differences between the information contained in the questions you were asked earlier and the information in the video”, Higham, 1998, p. 271, see also Greene et al., 1982). For example, in the experiment by Lindsay (1990) a strong warning (“there is no question in this test for which the correct answer was mentioned in the story”; Lindsay, 1990, p.1080) was effective. Wright (1993) provided an even stronger, more direct warning (“The narrative you have just read had one incorrect fact. In the slide sequence the woman did NOT have any cereal with her breakfast”, Wright, 1993) and while this was effective, some participants still ‘remembered’ the woman having cereal with her breakfast. On the other hand, Eakin, Schreiber, and Sergent-Marshall (2003) found both a general warning, warning participants that some details in the narrative may be inaccurate, and a specific warning, mentioning specifically what details were inaccurate, reduced the misinformation effect. Another aspect of providing a warning that may reduce participants’ susceptibility to misinformation is when there is an element of social influence involved.

When eyewitnesses are subject to the social influence of others, they are less likely to report misinformation (Chambers & Zaragoza, 2001). For instance, when participants are urged by the experimenter to present themselves in a positive light, they are able to resist reporting the misinformation (see Blank, 1998, p. 492; Zaragoza & Koshmider, 1989, p. 255). Participants are also able to resist misinformation when the source of the post-event information is revealed to be a person who is presenting the information in a biased way (Echterhoff, Hirst, & Hussy, 2005). Research on psychological reactance (an aversive affective reaction that occurs when free behaviour is restricted, Brehm, 1966) has shown that people react to attempts of others to control or determine their actions by applying internal control over their thoughts or actions (Echterhoff, Hirst, & Hussy, 2005). Thus, when it is revealed to an eyewitness that the person summarising the details of the event may not be completely accurate, they are able to resist reporting the misinformation. This way of reducing the influence of misinformation is termed a social post-warning by Echterhoff, Hirst, and Hussy (2005).

While several studies have examined the ability of social post-warnings to reduce the misinformation effect (eg., Chambers & Zaragoza, 2001; Greene et al., 1982; Highhouse & Bottrill, 1995; Meade & Roediger, 2002, Experiment 1), Echterhoff, Hirst, and Hussy (2005) go a step further by separating the effects of an explicit post-warning and a social post-warning. Interestingly, the studies that involved both an explicit post-warning and a social post-warning (Chambers & Zaragoza, 2001; Greene et al., 1982; Highhouse & Bottrill, 1995; Meade & Roediger, 2002, Experiment 1) all used a weak form of explicit post-warning, that in previous studies have been found unsuccessful in reducing the misinformation effect (see Belli et al., 1994; Christiaansen & Ochalek, 1983; Frost et al., 2002; Greene et al., 1982; Higham, 1998; Zaragoza & Lane, 1994). However, three of the studies found significant decreases in the amount of misinformation reported (Chambers & Zaragoza, 2001; Highhouse & Bottrill, 1995; Meade & Roediger, 2002, Experiment 1), suggesting that the social post-warning plays a significant role in reducing the misinformation effect (Echterhoff, Hirst, & Hussy, 2005).

Echterhoff, Hirst, and Hussy (2005) carried out four experiments looking at the effect of different types of social post-warnings (characterising the source as either trustworthy or incompetent), as well as different types of memory tests (cued-recall and recognition) and found that post-warnings reduced the misinformation effect in all four experiments. Thus, the effect of a post-warning in reducing the misinformation effect is robust.

So how do warnings reduce the misinformation effect? With a pre-warning, the answer is clear. According to Craik and Lockhart (1972), a pre-warning can serve to reduce participants' depth of encoding of the material. That is, after being warned that the information they are about to receive may be biased, they can encode the material on a shallow, rather than deep, level. Both explicit pre-warnings (Greene et al., 1982) and social pre-warnings (Dodd & Bradshaw, 1980; Smith & Ellsworth, 1987; Underwood & Pezdek, 1998) have been found to reduce the misinformation effect. Of course, as discussed above, in most real-life cases a pre-warning is not feasible. While a difference in the level of encoding on the post-event information can explain the reduction in the misinformation effect following a pre-warning, we know that this encoding difference is not present in the case of a post-warning. If participants receive a warning only after they have already received the information, they will have already encoded it, although it is not clear whether this information is encoded as a separate memory or assimilated into the existing memory for the original event. Thus, something must happen after the post-warning that results in a reduction of the misinformation recalled. Thomas et al. (2010) propose that when warned, participants then switch from responding based on retrieval fluency, to engaging in a more thoughtful and effortful evaluation of the information that comes to mind. Thomas et al. (2010) termed this the retrieval fluency hypothesis, and a discussion of this follows.

Retrieval Fluency

Thomas et al. (2010) used the retrieval fluency hypothesis to explain the RES effect. They proposed that when participants in the typical RES procedure read the post-event information containing misinformation after they have taken the initial test, their attention is drawn to the details in the post-event information that correspond to details that are the focus of the questions in the test. Thus, these particular details are preferentially encoded, and therefore are more easily retrieved later, in the final test. Ease of retrieval, or retrieval fluency, has been found to be related with a persons' judgement of their confidence in the accuracy of recall (Kelley & Lindsay, 1993; Koriat, 1993; Koriat, Ma'ayan, & Nussinson, 2006).

The link between confidence and retrieval latency is central to Thomas et al.'s (2010) retrieval fluency hypothesis. When we talk about a persons' confidence, we are talking about the metacognitive judgement they make about whether their response or answer is correct or incorrect (Ackerman & Koriat, 2011). It has been found that people tend to base their confidence judgements on how quickly the answer came to mind. For instance, when an answer springs to mind quickly, people assume it must be correct, and thus are more confident about the accuracy of their answer. Conversely, if it takes a longer time to think of the answer, people are more likely to assume the answer is incorrect, and thus be less confident in the accuracy of their answer. The ease with which an answer comes to mind was termed 'fluency' by Jacoby, who researched the subjective experience of remembering (see, for example, Kelley & Jacoby, 1990). To examine this phenomenon, Jacoby and colleagues carried out experiments which facilitated participants' processing of distractor items, which resulted in them later falsely 'remembering' those items (Jacoby & Whitehouse, 1989; Kelly, Jacoby, & Hollingshead, 1989; Lindsay & Kelley, 1991, cited in Kelley & Lindsay, 1993; Whittlesea, Jacoby, & Girard, 1990). It was found that when participants unconsciously processed information, this led to a feeling of familiarity, which was referred to as an 'illusion of memory' (Jacoby & Whitehouse, 1989, p. 132). Thus, when participants are tested, they tend to see the items they recognise, and because these are easily retrieved from memory the participants have a subjective experience of remembering those items. The idea of the unconscious processing of information leading to a feeling of familiarity is akin to Tversky and Kahneman's (1973) availability heuristic, which is the tendency to make judgements of probabilities or frequencies based on the ease with which they can think of examples.

Kelley and Lindsay (1993) took this idea one step further. They argued that when an answer comes to mind quickly, this leads to a subjective experience of remembering, which results in an increased confidence in the accuracy of the answer – regardless of whether the answer is correct or incorrect. This point is important, because the mere fact that retrieval latency is related to confidence is not astonishing in itself – it is to be expected that well-known answers would be retrieved more rapidly than lesser-known answers, and also that people will be highly confident about well-known answers. To investigate whether ease of retrieval plays a causal role in peoples' confidence, Kelley and Lindsay (1993) manipulated ease of retrieval to see what effect this would have on participants' confidence in their answers. Kelley and Lindsay (1993) found that confidence was related to retrieval fluency in an experiment where they exposed participants to correct and related-but-incorrect answers to general knowledge questions. When participants were exposed to the answers, whether they were correct or related-but-incorrect, they came to mind quickly and easily in a later test, and thus participants were more confident in their answers (Kelley & Lindsay, 1993). The important aspect of this study is that it showed that when participants are exposed to information, in this case the answers to general knowledge questions, the retrieval fluency does not depend on whether the answer is correct or incorrect. When participants are exposed to information, this later comes to mind easily, and they are therefore more confident in the accuracy of the answer than participants who were not exposed to the information.

The negative correlation between retrieval latency and confidence in the accuracy of the answer has been found by other authors also. Nelson and Narens (1990) found a negative correlation between response latency and confidence ratings, and this relationship held whether participants' answers were correct or incorrect. Shaw (1996) lends further evidence for a negative correlation between confidence and retrieval latency. Participants answered forced-choice questions after viewing a simulated crime scene, and were then questioned about their responses. This post-event questioning led to increased confidence in their later answers, even when the answers were incorrect. Shaw (1996) claims that this effect is explained by the retrieval fluency hypothesis.

Studies such as the ones mentioned above lend weight to the theory that in forming metacognitive judgements, such as the judgement of an answer to be accurate, people rely on feedback from task performance (Hoffmann-Biencourt, Lockl, Schneider, Ackerman, & Koriat, 2010; Koriat & Ackerman, 2010a; Koriat et al., 2006; Koriat, Ackerman, Lockl, & Schneider, 2009a, 2009b). When people perform a task, mnemonic cues from their task performance bring about metacognitive feelings in the person through a process which operates below the level of full consciousness (Kelley & Lindsay, 1993; Koriat & Ackerman, 2010b; Koriat & Levy-Sadot, 1999). In relation to the negative correlation between confidence and retrieval latency, the mnemonic cue people get from task performance is in the form of processing fluency, and this leads to metacognitive feelings of knowing and confidence. These metacognitive feelings of knowing and confidence lead people to think that their answer is accurate.

In terms of the RES effect, the above findings are very important. As stated by Thomas et al. (2010), the negative relationship between confidence and retrieval latency leads participants in the repeated testing condition to believe the misleading details in the post-event information are accurate, as they come to mind quickly. This is what leads them to endorse more of the misinformation items on the final test, compared to participants in the single testing condition whose attention is not as strongly drawn to the details in the post-event information, and therefore the misinformation items do not come to mind more quickly than correct items.

Summary and Hypotheses

The current study aims to further investigate the RES effect, particularly with respect to the retrieval fluency hypothesis proposed by Thomas et al. (2010). To recapitulate, Thomas et al. (2010) claim that the RES effect can be explained by retrieval fluency-based responding. That is, participants who take an initial test and then are presented with post-event information containing details that relate to information in the initial test questions have their attention drawn to such details, and they are therefore better encoded. When it comes to the final test, participants respond based on retrieval fluency, or how quickly the answers come to mind. Because the post-event information enhances encoding for both the consistent and the misinformation items, participants are more likely to endorse these items on the final test (Thomas et al., 2010). The current study aims to partially replicate the Thomas et al. (2010) study, by investigating whether the RES effect will be found and whether it can be reduced by a warning. The current study will also investigate how robust the RES effect by using a different witnessed event and a shorter distraction phase. Previous studies have used the same 40 minute episode of the television programme “24” (Chan et al., 2009; Thomas et al., 2010; Chan & Langley, 2011; Chan & LaPaglia, 2011). The current study utilises a 10 minute long action-packed video clip depicting a bank robbery, which is more akin to the type and length of even an eyewitness may be exposed to in a real-life situation. Further, the current study uses a distraction phase of only two minutes, in comparison to the distractor phase that lasts 25 minutes in the standard RES procedure (e.g., Chan et al., 2009; Thomas et al., 2010). As the RES effect increases with the delay between the witnessed event and the presentation of misinformation (Chan & Langley, 2011), it is useful to find out whether the distractor phase in the current experiment will be long enough to induce an RES effect.

The current study focuses on young to middle-aged adults, between the ages of 18 and 45, in order to have a homogenous sample with respect to memory ability. It has been shown that young children do not have the same propensity to remember or report details from an event they have witnessed (see Goodman & Melinder, 2007, for a review). Also, older adults tend to be more susceptible to memory distortions and false recollections (Schacter, Koutstaal, & Norman, 1997), and it has been shown that warnings are not as effective with older adults as they are with young adults (Dehon & Bredart, 2004; Watson, McDermott, & Balota, 2004). Thus, we limited the current sample to people aged between 18 and 45 to minimise the variability in responses in order to better identify differences between the experimental groups.

Based on the findings of Thomas et al. (2010), it is expected that participants will respond to questions on the final test based on retrieval fluency. Thus, it is hypothesised that participants who take an initial test will endorse more of the consistent and misinformation items (the items that correspond to details mentioned in the post-event information) on the final test, compared with those who do not take an initial test. However, this effect will be reduced or eliminated for participants in the warning condition. It is also hypothesised that because participants respond based on retrieval fluency, participants who take an initial test will have faster response times for consistent and misinformation items, compared to participants in the single testing condition.

Again, this effect will be reduced or eliminated for participants in the warning condition. Further, it is hypothesised that when participants are warned, the reaction time for those who took an initial test will be slower than those in the single testing condition, as they will then switch from fluency-based responding to source monitoring-based responding. Finally, due to the negative correlation between retrieval latency and confidence, participants in the repeated testing condition will have higher confidence ratings for those answers that come to mind easily, ie., the consistent and the misinformation items in the post-event information. This effect will be reversed for those in the warning condition.

CHAPTER THREE

Method

The aim of the current research was to investigate whether the retrieval-enhanced suggestibility (RES) effect would be found using a shorter video event and a shorter distraction time, and whether the effect can be eliminated by warning participants that some of the information they received may not be accurate. The retrieval fluency hypothesis is that an initial test makes the information readily available, so that the aspects of the misinformation that relate to the initial test are more easily retrieved later on. This was be tested by examining confidence ratings and response latency measures. The study used an experiment with a mixed 2 x 2 x 3 factorial design, meaning that there were three independent variables of interest. The first independent variable, testing condition, had two levels (single testing and repeated testing), and was manipulated between subjects. The second independent variable, warning condition, also had two levels (no warning and warning), and was also manipulated between subjects. The third independent variable was item type and this had three levels (control, consistent, and misinformation). Item type was manipulated within subjects.

Participants

Participants were 91 people from across New Zealand. Participants were recruited via social media and word of mouth, and were required to be between 18 and 45 years of age and to be proficient in English and the use of the computer. Participants were volunteers and no compensation was given. Participants were assigned to the experimental conditions automatically via the computer software. The software served up the first version (experimental condition) until that version was completed, and then it served up the next version and the process continued. As there was some imbalance towards the end of the data collection phase, the version for the last few participants was set manually to ensure a relatively equal number of participants across the four experimental conditions. The participant group was made up of 64 (70.3%) females and 27 (29.7%) males, and the ethnicity distribution was 77% New Zealand European, 4% Asian, 3% Maori, 2% Pacific Peoples, 1% Middle Eastern/Latin American/African and 4% Other. The age distribution was as follows: 51.6% between 18-25, 16.5% between 26-30, 11.0% between 31-35, 9.9% between 36-40, 11.0% between 41-45.

Materials and Procedure

Participants were invited to complete the experiment online. They were instructed to carry out the experiment somewhere quiet, where they were alone, using either a desktop computer, a laptop computer, or a tablet. Participants were asked to read an information sheet (provided in Appendix A) explaining what the study involved, and were informed that continuing with the experiment was deemed to be consent.

The experiment used HTML, JavaScript and CSS software, with the videos hosted on YouTube and the Tetris task embedded as a Flash object. When the participant agreed to participate, the program was uploaded onto their browser until they completed the experiment, at which point the data was passed back to the server.

Participants watched a video excerpt from the film *44 minutes,* which depicted a bank robbery known as the North Hollywood shootout. The excerpt was 9 minutes 40 seconds long. Following the video, participants in the single testing condition completed a 15 question multi-choice test. Each question asked about a detail in the event, giving a forced-choice between four plausible alternative answers. For example, one of the questions was "What type of vehicle do the bank robbers drive?", with the option to answer van, SUV, sedan or station wagon. All questions are included in Appendix B. After each question, participants were instructed to indicate their confidence in that answer, on a scale between 0% to 100%. Meanwhile, participants in the repeated testing condition played the video game Tetris for three minutes, approximately the same time it took for participants in the test condition to complete the test.

After either completing the test or playing Tetris, all participants watched a short video clip about the change-blindness effect as a distraction task. The video lasted for approximately two minutes. After watching the distractor video all participants were then instructed to read through a précis of the *44 Minutes* clip. The précis was adapted from the information about the film and the real event found on the website *Wikipedia* (n.d.). Embedded in the précis were 15 critical details about the film. Each critical detail was matched to the content of one of the 15 questions in the memory test, and only appeared once in the précis. Of the 15 critical details, five were consistent with the film, five were misleading (they were inconsistent with the film), and five were control details (they did not feature in the summary). For example, a misleading detail for the question "What are the two police officer's talking about while patrolling the area, before they notice the bank robbers?" would be "two police officers are doing their rounds, talking about their weekend", while a consistent detail would be "two police officers are doing their rounds, talking about what they had for breakfast". The control detail would be "two police officers are doing their rounds". This task took approximately three minutes.

Following the misinformation task, participants filled out a short demographic questionnaire which gathered information about the participants' age, gender, and ethnicity, and also asked whether they had seen the film *44 minutes* before.

Finally, all participants took a memory test, which was exactly the same as the participants in the test group had taken earlier. Participants were then thanked for taking part in the experiment, invited to leave their contact details if they wanted to receive a summary of the results of the experiment, and debriefed.

CHAPTER FOUR

Results

Introduction

As discussed in previous chapters, the current research focused on the retrieval-enhanced suggestibility (RES) effect, in particular the retrieval fluency hypothesis in relation to the RES effect, across young to middle-aged adults. The data were analysed using the software programme Statistical Package for the Social Sciences (SPSS; version 22.0).

Initial analyses

As the use of analysis of variance (ANOVA) requires certain assumptions to be met, some preliminary analyses were carried out to ensure that ANOVA would be a suitable technique to use, given the data. The use of ANOVA requires that the dependent variables are approximately normally distributed, and variances are homogenous (Pallant, 2013). In addition to these assumptions, the mixed within-between subjects ANOVA requires the assumption of homogeneity of intercorrelations to be met (Pallant, 2013).

Normality. Normality of each dependent variable was assessed in a number of ways. Normality was tested using the Shapiro-Wilk statistic, as this is considered more appropriate for small to medium sample sizes (Allen & Bennett, 2012). Normality was also assessed by examining skewness and kurtosis values, and viewing Q-Q plots and histograms.

Average hit rate. The average hit rate is the proportion of the questions on the final test that participants answered correctly. Each question on the test contained information that matched either a control detail, a consistent detail, or a misinformation detail in the post-event information, thus the test questions are classified as either control, consistent or misinformation. As there are five questions for each item type, the number of correct responses was divided by five to find the average accurate hit rate. Results from the Kolmogorov-Smirnov tests indicated that the average hit rate was not normally distributed for all three items types, the control items $D(91) = 0.87$, $p < .001$, the misinformation items $D(91) = 0.93$, $p < .001$, and the consistent items $D(91) = 0.77$, $p < .001$. Further, average hit rate scores were significantly negatively skewed for control ($z_{skewness} = -2.13$, $p = .05$) and consistent items ($z_{skewness} = -4.98$, $p = .05$). The histogram and Q-Q plots confirm that control and consistent scores were not normally distributed (see Appendix C). Because the data for the average hit rate violates the assumption of normality, a requirement of ANOVA, and there was no appropriate transformation that could be applied to the data, the results from the analyses of the average hit rate data were interpreted with caution.

Average false recognition of misinformation. For the questions relating to misleading information, participants could choose either the misleading lure, the correct answer, or one of two plausible alternatives. The average false recognition of misinformation score is the proportion of misleading lures that were chosen.

Results from the Kolmogorov-Smirnov tests indicated that the average false recognition of misinformation was not normally distributed, $D(91) = 0.91$, $p < .001$. The scores were significantly positively skewed (zskewness = 2.18, $p = .05$). The histogram showed the scores are positively skewed, and the Q-Q plot shows a non-linear trend (see Appendix D). Thus, it is clear the scores violate the assumption of normality. The scores for the false recognition of misinformation rate were transformed using a square root transformation. The resulting distribution was approximately normally distributed.

Average confidence ratings. After choosing their answer, participants then had to rate their confidence in that answer using a number between 0-100. The average confidence scores were then calculated for each item type.

Average confidence scores were not normally distributed, as indicated by the Kolmogorov-Smirnov scores (control items: $D(91) = 0.96$, $p = .01$, misleading items: $D(91) = 0.95$, $p < .001$, consistent items: $D(91) = 0.94$, $p < .001$). The scores were significantly negatively skewed for control items (zskewness = -2.34, $p = .05$), consistent items (zskewness = -2.76, $p = .05$), and misinformation items (zskewness = -3.11, $p = .05$). After viewing the histograms and Q-Q plots (see Appendix E), it was concluded that the average confidence scores were not normally distributed.

After applying a reflect and square root transformation, the average confidence scores were approximately normally distributed. Thus, the transformed values for average confidence scores were used in the analysis.

Average confidence ratings for misleading questions. For questions where the corresponding detail in the post-event information was misleading, we calculated an average confidence score for questions where participants chose the correct answer, and an average confidence score for questions where participants chose the misleading lure.

Average confidence for misleading trials scores were not normally distributed (correct items: $D(91) = .92$, $p < .001$, misinformation items: $D(91) = .91$, $p < .001$), and were negatively skewed (correct items ($z_{skewness} = -3.07$, p = .05), misinformation items ($z_{skewness} = -2.15$, p = .05). Kurtosis scores were significant for misleading trials only ($z_{kurtosis} = 2.77$, p = .05). The histogram and Q-Q plots confirmed that scores were not normally distributed (see Appendix F).

Based on the appearance of the histogram plots, and the significant results of the Kolmogorov-Smirnov and skewness tests, a reflect and square root transformation was applied. After transforming the variables, scores were approximately normally distributed.

Average reaction time scores*.* The time from when participants were given each question to when they made a selection from the four available options was measured. The average reaction times were then calculated for each item type.

The average RT scores for control and consistent items contained extreme outliers, as evident in their respective boxplots (see Appendix G). Extreme outliers are data points that extend more than three box lengths from the edge of the box (Pallant, 2013). These outliers were removed.

The average reaction time scores did not meet the assumption of a normal distribution. Results from the Kolmogorov-Smirnov tests indicated that the average reaction time deviated from a normal distribution for all three items types, the control items (*D*(91) = .92, $p < .001$), the misinformation items (*D*(91) = 0.92 $p < .001$), and the consistent items (*D*(91) = .90, $p < .001$). Further, all item types were positively skewed (control items ($z_{skewness} = 6.75$, $p = .05$), consistent items ($z_{skewness} = 7.37$, $p = .05$), and misinformation items ($z_{skewness} = 3.27$, $p = .05$). The histogram and Q-Q plots confirmed that average reaction time scores were not normally distributed (see Appendix G). After applying a square root transformation, reaction time scores for all three item types appeared approximately normally distributed on viewing the respective histogram plots. Thus, the transformed values for average confidence scores were used in the analysis.

Average reaction time scores for misleading trials. Like the average confidence rating for misleading questions score, we calculated an average reaction time score for questions where participants chose the correct answer, and an average reaction time score for questions where participants chose the misleading lure.

Average reaction time scores for misleading questions did not meet the assumption of normality. Scores deviated from a normal distribution for both correct items (*D*(91) = .93, $p < .001$) and misinformation items (*D*(91) = .89, *p* = < .001), and both item types were positively skewed (correct items ($z_{skewness}$ = -2.69, $p = .05$), misinformation items ($z_{skewness} = 4.60$, $p = .05$). The histogram and Q-Q plots also showed that scores were not normally distributed (see Appendix I).

Based on the appearance of the histogram plots, and the significant results of the Kolmogorov-Smirnov and skewness tests, a square root transformation was applied. The transformed data was approximately normally distributed.

Homogeneity of variance. The use of ANOVA requires that the samples are obtained from populations of equal variances (Pallant, 2013). This assumption was assessed using Levene's test for equality of variances. Violation of the homogeneity of variances assumption is not particularly concerning because ANOVA is reasonably robust to violations in this assumption, provided the group sizes are of a reasonably similar size (Stevens, 2009). In the present study, the group sizes are fairly similar so we do not need to be too concerned about data that violate this assumption. The assumption of homogeneity of variance will be discussed in the appropriate sections only if the assumption is violated.

Homogeneity of intercorrelations. For mixed between-within subjects ANOVA, there is an additional assumption that must be met. This is the assumption that for each level of the between-subjects variable, the pattern of intercorrelations among the levels of the within-subjects variable should be the same (Pallant, 2013). This assumption is tested using Box's M statistic. As with the homogeneity of variance assumption, violations to the homogeneity of intercorrelations assumption will be discussed in the appropriate section. Otherwise, this assumption has been met.

Main Findings

Recognition accuracy.

Table 4.1

Average hit rates on the final test as a function of test condition, warning, and item type (standard deviations are in parentheses).

	Consistent	Control	Misinformation
No warning			
Single testing	.87 (.14)	.73 (.21)	.44 (.22)
Repeated Testing	.89 (.17)	.75 (.20)	.45 (.32)
Warning			
Single testing	.86 (.17)	.49 (.25)	.49 (.25)
Repeated Testing	.82 (.29)	.81 (.19)	.60 (.27)

Based on the RES effect, we expected average hit rates for misinformation items to be lower after repeated testing, as testing increases susceptibility to misinformation. After a warning, however, we expected this effect to be attenuated. We also expected that the average hit rate for consistent items would be higher for repeated testing, compared to single testing, and this effect would also be attenuated with a warning. We expected neither warning nor repeated testing to have an effect on the average hit rates for control items. The descriptive statistics given in Table 4.1 shows that for consistent items, participants in the repeated testing conditions who were not warned had a higher average hit rate than participants in the other conditions, consistent with the RES effect. Participants in the warning conditions had slightly lower average hit rates, which is contrary to the RES effect. The average hit rate for control items was lower for participants who were warned and were in the single testing condition, but higher for those who were warned and were in the repeated testing condition. This is inconsistent with RES as testing should have no effect of the average hit rate for control items for participants who had been warned about the potential inaccuracy of the information. For the misinformation items, there was an increase in average hit rate with both testing and warning, contrary to predictions. For both warning conditions, repeated testing increased average hit rate slightly compared to single testing, providing evidence for RES. Average hit rates were highest for consistent items and lowest for misinformation items. Overall, there was limited evidence for the RES effect, and in fact warning participants seemed to result in a lower average hit rate, contrary to predictions.

Because many participants had high average hit rates for consistent and control items and few participants had low average hit rates, the data were negatively skewed. As the data for misinformation items were slightly positively skewed, it was not possible to find a transformation that could be applied to scores from all three item types that would make the distributions approximately normal. As there is no non-parametric equivalent to the mixed design ANOVA, an ANOVA was carried out on the average hit rate data. The results from ANOVA analyses are fairly robust to minor violations to the assumption of normality with reasonable sized samples (Stevens, 2009). However, as our group sizes are relatively small (22+), it cannot be assumed that the ANOVA will be robust to the violation of the normality assumption, and therefore the results must be interpreted with caution.

A 3 (item type) x 2 (test condition) x 2 (warning condition) mixed model ANOVA was conducted on the average hit rate data. There was a significant interaction between item type and warning condition (Wilk's Lambda = .93, F (2, 86) = 3.07, p = .05, partial eta squared = .07). The main effect of item type was significant (Wilk's Lambda = .38, F (2, 86) = 69.48, p < .001, partial eta squared = .62). Pairwise comparisons indicated that the average hit rate for misinformation items was lower than both the control items (p < .001) and the consistent items (p <.001). Further, the average hit rate for consistent items was higher than for control items (p = .001). The three-way interaction was not significant, and neither was the interaction between item type and test condition. The main effect of warning condition was also not significant.

To further examine the interaction between item type and warning condition in the 3 x 2 x 2 ANOVA, a simple effects analysis was conducted examining the effect of warning condition on each item type. For the misinformation items, the main effect of warning nearly reached significance (F (1, 89) = 3.44, p = .07, partial eta squared = .04), such that participants who were warned had higher (M = .55, SD = 0.4) average hit rates than participants who were not warned (M = .44, SD = .04). For control (F (1, 89) = 2.44, p = .12, partial eta squared = .03) and consistent (F (1, 89) = 1.45, p = .23, partial eta squared = .02) items, the main effect of warning was not significant. These results indicate that warning participants results in more accurate responses to misinformation items, but has no effect on the average hit rates for control and consistent items. However, as the effect was not statistically significant, we cannot conclude that an RES effect was found in the current study.

False recognition associated with misleading questions.

Table 4.2

Average false recognition of misinformation associated with misleading questions (standard deviations in parentheses)

	No Warning	Warning
Single Testing	.34 (.25)	.33 (.23)
Repeated Testing	.43 (.32)	.30 (.28)

Based on the descriptive statistics presented in Table 4.2, it appears that for single testing participants, those who were not warned about the potential inaccuracy of the post-event information had a higher average false recognition rate than those who were warned, meaning that those who were warned were more likely to choose the misleading lure than those who were not warned. This indicates that taking an initial test increases susceptibility to the RES effect, but this is reversed with a warning, which is consistent with RES. A 2 (test condition) x 2 (warning condition) between subjects ANOVA was conducted on the false recognition of misinformation to see whether this apparent effect was statistically significant. Neither the main effect for test condition (F (1, 87) = .001, p = .98, partial eta squared < .001) or warning condition (F (1, 87) = 1.12, p = .29, partial eta squared = .01) were significant. The interaction between warning condition and test condition was also not significant (F (1, 87) = .77, p = .38, partial eta squared = .01).

Confidence ratings.

Table 4.3

Mean confidence (0-100) associated with the final test as a function of warning, test condition and item type (standard deviations in parentheses).

	Consistent	Control	Misinformation
No warning			
Single testing	77.99 (14.98)	80.88 (12.81)	72.73 (20.76)
Repeated Testing	88.24 (12.52)	79.06 (13.16)	77.39 (14.81
Warning			
Single testing	69.97 (15.71)	75.30 (14.42)	62.22 (21.04)
Repeated Testing	75.74 (20.75)	79.62 (18.03)	69.62 (21.33)

According to the descriptive statistics given in Table 4.3, for consistent items, there appears to have been an increase in confidence with repeated testing, consistent with the retrieval fluency hypothesis. For misinformation items, there was a decrease in confidence for the warning conditions, consistent with predictions. Neither warning nor testing seems to have had any effect on the average confidence for control items. Overall, confidence was slightly lower for misinformation items, compared to consistent and control items. For those who took an initial test and received a warning, confidence was lower for the consistent and misinformation items, as predicted by the retrieval fluency hypothesis. Thus, the descriptive statistics are weakly consistent with the retrieval fluency hypothesis.

A 3 (item type) x 2 (test condition) x 2 (warning condition) mixed model ANOVA was conducted on the confidence data to statistically test for an RES effect. According to the retrieval fluency hypothesis, we would expect confidence to be higher for misinformation and consistent items after repeated testing, but only in the no warning condition. Results from the ANOVA show there was a significant interaction between item type and warning condition (Wilks' Lambda = .91, F (2, 86) = 4.09, p = .02, partial eta squared = .09). The main effect of test condition was significant (F (1, 87) = 5.02, p = .03, partial eta squared = .06). Confidence scores were higher for repeated testing participants (M = 78.58, SD = 2.03), compared to single testing participants (M = 72.58, SD = 2.00). The main effect of warning condition was also significant (F (1, 87) = 5.56, p = .02, partial eta squared = .06). Confidence scores were higher for participants who were not warned (M = 79.10, SD = 2.00) than for participants who were warned (M = 72.06, SD = 2.03). The three way interaction and the interaction between item type and test condition were not significant.

To investigate the interaction between item type and warning condition, a 3 (item type) x 2 (test condition) ANOVA was conducted for each warning condition separately. In the ANOVA examining only those in the no warning conditions, the homogeneity of variance assumption was met for control and consistent items, but not misinformation items. As discussed earlier in this chapter, violations to the homogeneity of variance assumption are not too concerning when sample sizes are equivalent, as they are in the current study. There was a significant main effect for item type (F (2, 43) = 5.21, p = .01, partial eta squared = .20). Pairwise comparisons indicated that the confidence for consistent items was significantly higher than for control items (p = .05) and misinformation items (p = .003). There was no statistical difference between confidence ratings of control and misinformation items. The main effect for test condition was also significant (F (1, 44) = 2.73, p = .05, partial eta squared = .09), such that single testing participants had lower confidence (M = 76.04, SD = 2.38) in their responses than repeated testing participants (M = 82.17, SD = 2.50). When analysed further, the main effect of test was only significant for consistent items, such that confidence was higher for repeated testing participants than single testing participants. The interaction between item type and test was significant (F (2, 44) = 3.28, p = .04, partial eta squared = .07). From the interaction plot, it appears that for consistent items only, confidence is higher for participants in the repeated testing condition, compared to those in the single testing condition. For both misinformation and control items, there was no difference in confidence between the single testing and repeated testing conditions. This provides some evidence for the retrieval fluency hypothesis, as it is expected that confidence would be higher for items that have been seen before (on the initial test) and therefore better encoded. However, it is expected that confidence would be higher for misinformation items also.

In the ANOVA examining only those in the warning conditions, there was a significant main effect for item type (F (2, 42) = 98.45, p < .001, partial eta squared = .82). People gave lower confidence ratings for misinformation items than consistent items (p = .02) or control items (p < .001), and lower confidence ratings for consistent items than control items (p = .05). The interaction between item type and test condition was not significant (F (2, 42) = .94, p = .95, partial eta squared = .04), and there was no main effect of test condition (F (1, 43) = 2.25, p = .13, partial eta squared = .05).

To further examine the interaction between item type and warning condition in the 3 x 2 x 2 ANOVA, a simple effects analysis was conducted examining the effects of warning condition on each item type. There was a significant main effect of warning for misinformation items (F (1, 89) = 4.97, p = .03, partial eta squared = .05), such that participants who were warned had lower (M = 65.92, SD = 2.94) confidence in the misinformation items than participants who were not warned (M = 75.06, SD = 2.91). There was also a significant main effect of warning for consistent items (F (1, 89) = 8.41, p = .01, partial eta squared = .09), such that participants who were warned had lower (M = 72.86, SD = 2.43) confidence in the consistent items than participants who were not warned (M = 83.11, SD = 2.41). However, for control items, there was no difference in confidence between participants in the warning and the no warning conditions (F (1, 89) = .03, p = .89, partial eta squared < .001). These results indicate that warning participants has the effect of lowering their confidence in their answers for both misinformation and consistent items (ie., questions for which details were provided in the post-event information).

Confidence on misleading trials.

Table 4.4

Mean confidence scores (0-100) associated with recognition decisions for misleading trials (standard deviations in parentheses).

	Correct	Misinformation
No warning		
Single testing	76.71 (21.71)	62.61 (29.60)
Repeated Testing	75.54 (21.57)	76.56 (27.07)
Warning		
Single testing	62.19 (25.77)	56.51 (26.38)
Repeated Testing	74.74 (26.58)	51.41 (38.36)

The descriptive statistics provided in Table 4.4 show that mean confidence scores on misleading trials were lowest for participants who were given a warning and then endorsed misinformation items on the final test, regardless of testing condition. Participants were also less confident when endorsing correct responses after being warned. This suggests that warning participants makes them less confident when responding with both correct answers and misinformation. For misinformation responses, confidence was higher for repeated testing participants than single testing participants when participants were not warned, but lower when participants were warned. A 2 (test condition) x 2 (warning condition) between subjects ANOVA was conducted on the mean confidence scores for misinformation responses to investigate this further. Results show that there was a main effect of warning

(*F* (1, 69) = 5.30, *p* = .02, partial eta squared = .07), with participants who did not receive a warning (*M* = 69.58, *SD* = 5.00) having higher confidence in their answers than participants who did receive a warning (*M* = 53.96, *SD* = 5.17). The main effect of test condition was not significant, nor was the interaction.

To investigate whether the difference in confidence ratings for each warning condition depends on the test condition in the 2 x 2 ANOVA on confidence in responses to misinformation items, a simple effects analysis was conducted. In the ANOVA examining only those in the single testing conditions, there was no main effect of warning condition (*F* (1, 36) = .82, *p* = .37, partial eta squared = .02), showing that there was no significant difference in confidence ratings between participants who were warned and participants who were not warned. However, in the repeated testing condition, there was a main effect of warning (*F* (1, 33) = 4.59, *p* = .04, partial eta squared = .12), showing that participants who were not warned had higher (*M* = 76.56, *SD* = 27.07) confidence in their answers than participants who were warned (*M* = 51.41, *SD* = 38.36).

Finally, a 2 (test condition) x 2 (warning condition) between subjects ANOVA was also conducted for the mean confidence scores for correct responses to misleading questions. Neither the main effect of test condition (*F* (1, 84) = 1.94, *p* = .17, partial eta squared = .02) or warning condition (*F* (1, 84) = 1.86, *p* = .18, partial eta squared = .02) were significant.

Response latency.

Table 4.5

Mean response latencies (in ms) associated with the final test as a function of warning, test condition and item type (standard deviations in parentheses).

	Consistent	Control	Misinformation
No warning			
Single testing	7915 (2570)	5421 (2263)	6700 (2095)
Repeated Testing	4019 (1349)	3162 (1049)	3907 (2028)
Warning			
Single testing	9304 (3775)	5699 (1866)	7553 (2933)
Repeated Testing	4225 (1553)	3118 (1161)	4336 (2443)

As shown in Table 4.5, there is a clear trend across all item types showing that repeated testing participants responded faster than single testing participants. The lower response latency for repeated testing groups was predicted for consistent and misinformation items, but not for control items. Warning participants was predicted to reduce or eliminate the decrease in response latency, but there was, in fact, no effect of warning. Repeated testing also reduced variability among scores. Overall, results from response latencies suggest a clear effect of testing, but no effect of warning.

Results from a 3 (item type) x 2 (test condition) x 2 (warning condition) mixed between-within subjects ANOVA confirmed there was a significant interaction between item type and test condition (Wilks' Lambda = .82, F (2, 84) = 9.53, p < .001, partial eta squared = .19). The main effect of warning was not significant (F (1, 85) = 1.22, p = .27, partial eta squared = .01). There was no interaction between item type and warning condition, or test condition and warning condition. Levene's test of equality of error variances was significant for misinformation and control items, meaning that the assumption of homogeneity of variance was violated for consistent items only. Again, the violation of this assumption is not too concerning (see discussion earlier in the chapter).

To decompose the interaction between item type and test, a 3 (item type) x 2 (warning condition) ANOVA was conducted for each test condition separately. In the ANOVA examining only those in the single testing conditions, there was a main effect of item type (Wilks' Lambda = .34, F (2, 41) = 39.34, $p < .001$, partial eta squared = .66). Pairwise comparisons indicated that response latency for consistent items was significantly slower than for control items ($p < .001$) and misinformation items ($p < .001$). Further, response latency was significantly slower for misinformation items than for control items ($p < .001$). The interaction between item type and warning condition did not reach significance, and there was no main effect of warning. The assumption of equal variances was met for control and consistent items, but not for misinformation items.

In the ANOVA examining only those in the repeated testing conditions, there was also a main effect of item type (Wilks' Lambda = .44, F (2, 42) = 26.82, $p < .001$, partial eta squared = .56). Pairwise comparisons show that reaction times for control items were significantly faster than for both misinformation items ($p = .001$) and consistent items ($p < .001$). The difference in reaction times between misinformation and consistent items was not significant. The interaction between item type and warning was not significant, and there was no main effect of warning. Taken together, results from the ANOVAs examining response latencies for the test conditions separately show that after repeated testing, the difference in response latency between consistent and misinformation items is eliminated. This suggests that repeated testing increases the fluency of responding for misinformation items, as predicted by the retrieval fluency hypothesis.

Response latency for misleading trials.

Table 4.6

Mean response latencies (in ms) associated with recognition decisions on misleading trials (standard deviations in parentheses).

	Correct	Misinformation
No warning		
Single testing	7593 (3230)	6112 (2789)
Repeated Testing	3948 (2791)	3117 (1064)
Warning		
Single testing	7942 (3212)	7148 (3523)
Repeated Testing	4987 (3149)	4499 (2777)

Table 4.6 presents the descriptive statistics associated with misleading questions. These are questions that relate to details of the original event that were replaced with misleading details in the post-event information. Average response latencies are presented for correct responses (where participants chose the correct answer and were not deterred by the misleading details in the post-event information) and misinformation responses. Regardless of warning condition or response type, it is clear that repeated testing reduced

response time significantly. According to the retrieval fluency hypothesis, we would expect response times to be similar for correct and misinformation responses for participants who are not warned, but for response times for misinformation responses to increase when participants are warned. We conducted a 2 (response type: correct or misinformation) x 2 (test) x 2 (warning) mixed model ANOVA for response times on misleading trials to examine this further. The assumptions of homogeneity of variances was violated for consistent and misinformation items. None of the interactions were significant, but there was a main effect of response type (Wilks' Lambda = .93, F (1, 66) = 5.19, p = .03, partial eta squared = .07). Participants responded significantly faster when they gave a misinformation response (M = 5218.86, SD = 327.56) than when they gave a correct response (M = 6117.34, SD = 373.53). The main effect for test condition was also significant (F (1, 66) = 27.71, p < .0005, partial eta squared = .30), showing that participants who took a repeated test (M = 4137.51, SD = 428.21) responded significantly faster than those who took a single test (M = 7189.70, SD = 393.50). The main effect for warning did not reach significance.

Separate 2 (response type) x 2 (test condition) ANOVAs were conducted for each warning condition. The ANOVA analysing only those in the no warning condition found that the main effect of response type was nearly significant (Wilk's Lambda = .91, F (1, 34) = 3.52, p = .07, partial eta squared = .09), such that participants responded faster when giving a misinformation response (M = 4614.17, SD = 369.22), compared to when giving a correct response (M = 5770.33, SD = 510.44). The ANOVA analysing only those in the warning condition did not find a main effect of response type, (Wilk's Lambda = .96, F (1, 32) = 1.52, p = .23, partial eta squared = .05).

Summary

The results show mixed evidence for the RES effect and the retrieval fluency hypothesis. Evidence from the analyses of the average hit rate data showed that participants were more accurate on consistent items and less accurate on misinformation items, but this did not differ according to test condition, as would be expected by the retrieval fluency hypothesis. Participants who were warned were more accurate on misinformation items than participants who were not warned, although this was not significant. Results from an analysis of the misinformation responses showed that participants who were warned were more accurate on misleading trials, consistent with RES. Results from the analyses of confidence ratings show that overall, confidence is higher for participants in the repeated testing condition. For both misinformation and consistent items, but not control items, confidence was lower for participants who had been warned about the potential inaccuracy of information. The response latency results showed that there was an interaction between response type and test condition. In the single testing condition, participants responded fastest to control questions, and slowest to consistent questions. Similarly, in the repeated testing condition participants responded faster to control questions than both misinformation and consistent questions. Results from the analyses of the response latency associated with misinformation questions indicated that reaction times were faster when giving a misinformation response than when giving a correct response. Results also showed that participants who took a repeated test responded faster than those who took a single test, consistent with what was predicted by the retrieval fluency hypothesis. However, participants in the repeated testing conditions would be expected to respond faster on the final test, as they have previously read the questions. This limitation of the response latency data is discussed further in Chapter Five.

CHAPTER FIVE

Discussion

The purpose of the current study was to investigate whether the retrieval-enhanced suggestibility (RES) effect would be found using a shorter video event and a shorter distraction phase, and whether it could be reduced by warning participants about the potential inaccuracy of the post-event information. To do this, an experiment was carried out similar to that used by Chan and colleagues (e.g., Chan et al., 2009). We also measured participants' confidence and response latencies, in order to test the retrieval fluency hypothesis, as proposed by Thomas et al. (2010). According to the retrieval fluency hypothesis, participants in the repeated testing condition (who take a test before reading the post-event information) have their attention drawn to the details in the post-event information that relate to the questions on the test (misinformation and consistent items), and these details are preferentially encoded. Thus, when it comes to the final test, the misinformation and consistent items come to mind quicker for participants who took an initial test, leading them to endorse more of the pre-tested items on the final test, respond more quickly to pre-tested items, and be more confident in these items. To recapitulate, it was hypothesised that repeated testing would lead to an increased number of the misinformation and consistent items being chosen on the final test compared to single testing. Repeated testing would also lead to increased confidence for responses to consistent and misinformation items, and a faster response latency on consistent and misinformation items, compared to control items. However, when participants are warned, the above hypothesised effects were expected to be decreased or eliminated.

This chapter examines the findings of the current study in relation to previous research on the RES effect and the retrieval fluency hypothesis, and then discusses the limitations of the current research and recommendations for future research.

Discussion of Main Findings.

Recognition accuracy/RES effect findings. Unfortunately, in the current study the analyses of recognition accuracy data were problematic as the data were not normally distributed. As the sample sizes are relatively small, this may have had an effect on conclusions drawn from the current data (see discussion in Chapter Four). It was hypothesised that the initial test would influence participants' processing of the post-event information they were later provided with. Specifically, they would pay more attention to the consistent and misinformation details in the post-event information, as these would 'pop out' due to them being related to the questions on the initial test. Thus, the misinformation and consistent details would be better encoded, and participants who took an initial test would be more likely to endorse the consistent and misinformation responses on the final test. It was further hypothesised that when participants were warned, they would no longer respond based on retrieval fluency, but would engage in a more effortful recollection process that would outshine the fluency-related processing of the items, and therefore the RES effect would be eliminated. Results showed that participants were more accurate for consistent items than control and misinformation items, and more accurate on control items than misinformation items. This is consistent with the retrieval fluency explanation for the RES effect, as the consistent and misinformation items were better encoded and therefore participants endorsed more of the consistent and misinformation items on the final test. This led to higher accuracy for consistent items, and lower accuracy for misinformation items, as participants endorsed more of the misleading lures rather than the correct answers. Further analyses found that for misinformation items, accuracy was higher for participants who received a warning than participants who did not receive a warning, although this effect did not reach statistical significance. For both control and consistent items, there was no effect of warning condition. These results provide evidence for the retrieval fluency explanation for the RES effect. The higher accuracy for misinformation items for participants who were warned, compared to participants who were not warned, was expected, as after being given a warning participants must try to distinguish which is the correct answer from the two sources of information they have been given. That is, they no longer respond based on which answer comes to mind easily (the misleading lure) but engage in more effortful recollection, which is more likely to lead to the correct answer. Thus, the results from the analysis of recognition data provides evidence, although not statistically significant, for retrieval fluency-based responding.

The analysis from the false recognition rates on misleading trials showed that participants endorsed more of the misinformation items after taking an initial test, but only when not warned. Although this effect was not significant, it was in the expected direction. Again, this shows that testing enhanced encoding of the misinformation items, leading these answers to come to mind more quickly and therefore be assumed to be correct. Thus, the evidence from the recognition accuracy data is consistent with the idea that people respond based on retrieval fluency, but only when they do not receive a warning about the potential inaccuracy of misinformation.

In summary, there is some evidence that suggests there was a RES effect in the current study, and that the RES effect can be explained by the retrieval fluency hypothesis. However, the number of consistent items endorsed on the final test was not higher for participants who were warned compared to participants who were not warned, as would have been expected. Further, the increase in the number of misinformation lures chosen on the final test by participants in the repeated testing condition compared to participants in the single testing condition was not significant, and while warning reduced the number of misinformation lures chosen by participants in the repeated testing condition, this also did not reach significance.

Generally, studies that have utilised the basic RES procedure have found an RES effect. That is, repeated testing has increased participants' susceptibility to misinformation. This has been the case in Chan and colleagues' (2009) study using a cued-recall test with both young adults and older adults. Chan et al. (2009) also found an RES effect using a modified-modified free-recall test, in which participants were invited to recall everything they could think of about the given cue. Thomas et al. (2010) found an RES effect in both their cued-recall and forced-choice recognition tests. Further, Chan and Langley (2011) found the RES effect in both of their two experiments. In the first, there was a 7-day delay between the post-event information and the final cued-recall test. In the second, there was a 7-day delay between the initial test, and the post-event information and final cued-recall test. Chan and Langley (2011) found that both participants who took an initial test and participants who only took the final test reported more misinformation in Experiment 2, when there was a 7-day delay between the initial test or distraction activity and the presentation of misinformation, than in Experiment 1 when there was a 7-day delay between the presentation of misinformation and the final test. This finding can be attributed to the recency effect, which is the tendency to have better recall for the most recently learned information than for information learnt earlier (Murdock, 1962). Thus, when participants were presented with the post-event information immediately before the final test, rather than seven days before the final test, they recalled more of the misleading details.

By contrast, in the four experiments carried out by Chan and LaPaglia (2011) it was found that while taking multiple initial tests increased participants' susceptibility to misinformation, an RES effect was not found in every experiment. In their first experiment, testing was manipulated between-subjects, with the number of initial tests taken varying between 0, 1, and 5. Recall of misinformation increased with the number of initial tests, with misinformation recall for one initial test being significantly higher than no initial test (the basic RES effect), but misinformation recall for five initial tests was not significantly higher than for one initial test. In their second experiment, testing was manipulated within-subjects. All participants took an initial test, but only half of the questions appeared on the initial test so that responses to tested and non-tested questions on the final test could be analysed. With the within-subjects testing, participants were more likely to report misinformation with a tested item, but the RES effect was only significant when participants had taken three or six initial tests. Clearly, the RES effect is not as robust when measured within-subjects. The final two experiments carried out by Chan and LaPaglia (2011) were modelled on the first two, with a seven-day delay being introduced before participants were presented with the post-event information and final test. After a seven-day delay, Experiment 3 found a higher level of misinformation recall for those taking one initial test compared to participants who did not take an initial test , although this effect did not reach statistical significance. Misinformation recall was significantly higher after taking five tests, compared to taking only one. Interestingly, in Experiment 3 testing also increased the number of correct responses to misinformation-related items. This is thought to be because initial testing enhanced both learning of the misinformation and retention of the original event. In Experiment 4, which varied the number of tests within-subjects and with a one-week delay, the RES effect was not significant. Consistent with Experiment 3, testing increased the number of correct responses to misinformation-related items. Experiment 4 also included a modified-modified free-recall (MMFR) test as the final test. The MMFR test instructs participants to recall everything they can that is associated with a given cue, regardless of the source of that information (for a more detailed explanation, see Chapter Two). There was an RES effect on the MMFR test. Chan and LaPaglia (2011) argue that the reason an RES effect was found in the MMFR test but not the cued-recall test was that participants had control over the information they reported in the cued-recall test, but in the MMFR test they could report multiple answers. When there were multiple sources of information (i.e., the witnessed event and the post-event information) participants must choose one answer to report on the cued-recall test, and this would be based on which answer is has a higher ‘memory strength’. Results from the MMFR test show that testing strengthened the memory for the original memory, therefore in the cued-recall test participants tended to report the answer with the strongest memory, the memory for the original event.

Chan and LaPaglia (2011) explain that the item-based, or within-subjects, RES effect is weaker than the participant-based, or between-subjects, RES effect is due to retrieval-induced facilitation. They argue that when taking an initial test, participants inadvertently recall target-*related* memories. That is, when answering a question participants may recall non-tested details that are related to, but not the same as, the target detail in the question. Chan and LaPaglia provide the following example to illustrate this concept: “when participants attempt to answer a questions about the main character’s shirt colour, they might spontaneously recall the shirt colour of his wife, with whom he was conversing” (2011, p. 427). As non-tested items may be covertly retrieved, this has the effect of making the tested and non-tested items more similar, thus weakening the within-subjects RES effect as compared to the between-subjects RES effect. As testing in the current experiment was manipulated between-subjects, however, we would expect to find an RES effect. Possible reasons for the lack of a statistically significant RES effect in the current study are discussed below.

The lack of a statistically significant RES effect in the current study may be explained by the very small delay between the initial test and the final test. In the standard RES procedure (e.g., Chan et al., 2009; Thomas et al., 2010) participants watch the video event, either take an initial test or undertake an unrelated task, and then take part in distractor tasks that last for approximately 25 minutes before receiving the post-event information and taking the final test. In comparison, in the current study there is a distractor phase that lasts for approximately three minutes, after taking the initial test and before receiving the post-event information and taking the final test. It is well documented that the longer the time interval between the witnessed event and the presentation of misinformation, the more likely eyewitnesses are to report the misinformation on a later test (see, for example, Loftus, Miller, & Burns, 1978). This is thought to be because as more time passes, eyewitnesses to an event are less likely to notice a discrepancy between the event details and the conflicting details presented in the post-event information. Research on the RES effect has also found that it increases with delay. Chan and Langley (2011) found that participants reported more misinformation when there was a 7-day delay between the initial test and the presentation of misinformation, compared to when there was a 7-day delay between the presentation of misinformation and the final test. Further, Chan and LaPaglia (2011) found that participants reported more misinformation after a 7-day delay than after the standard delay of approximately 25 minutes. Thus, perhaps in the current study the three minute distraction phase was not long enough for the memory trace to decay sufficiently, and thus for the RES effect to fully emerge.

Additionally, in each of the RES experiments carried out by Chan and colleagues, the same video has been used. Participants in these studies watched the same 40 minute episode of the television programme “24”. In contrast, the video in the current study was a 10 minute long action-packed video clip depicting a bank robbery. Given that there is a clear RES effect in the studies carried out by Chan and colleagues, but little statistically significant evidence for an RES effect in the current study, it may be useful for future research to look into how the witnessed event can effect whether people become more susceptible to misinformation after testing. In a real-life situation, it is unlikely that witnesses would have that length of time to witness the critical event. In that respect, the witnessed event in the current study is more akin to what eyewitnesses to real-life criminal events may experience.

Perhaps, in the current study, the combination of the briefer video event and the shorter delay combined to reduce the strength of the RES effect. According to the retrieval fluency hypothesis, the initial test questions serve as cues for the post-event misinformation presented later. Thus, repeated testing participants are expected to pay more attention to the misleading details in the post-event information, and preferentially encode these details. Perhaps it is the case that with a brief event and a short delay before the presentation of the post-event misinformation, repeated testing participants pay more attention to the misleading details, but do not preferentially encode the misleading details as they recognise them as misleading. That is, there is not enough time between the initial test and the presentation of misinformation for forgetting to set in. Future research will need to investigate whether the RES effect can occur with a brief event, and how long the delay needs to be between the initial test and the presentation of misinformation for the RES effect to occur.

Another possibility is that the current study may have failed to find significant results due to a lack of statistical power. According to Stevens (2009), when group sizes are small there is a possibility the non-significant results may be due to a lack of power rather than a lack of differences between the groups. The groups in the current study are relatively small, ranging from 22 to 24 participants in each of the four groups, therefore it is possible that the non-significant results were due to lack of power. The fact that the findings were in the expected direction given the hypothesis further suggests lack of power may be responsible. However, in the study by Chan et al. (2009), the groups were of a similar size to the groups in the current study, which indicates that the group sizes in the current study do allow for sufficient power to produce a significant RES effect.

Given that the results we were able to analyse provide only limited evidence for the RES effect, and non-significant results were in the expected direction, it is possible that the limited power of the current study, the shorter delay between the initial test, the post-event information and the final test, and the briefer video event contributed to the lack of significant evidence for an RES effect. Further research should be carried out examining whether eyewitnesses are more susceptible to misinformation after their memory has been tested in situations where there is little to no delay between the presentation of misinformation and the recall of that information.

Confidence findings. Based on the assumptions of the retrieval fluency hypothesis, we would expect confidence for responses to misinformation and consistent items to be higher for participants who take an initial test, but only when they are not warned about the potential inaccuracy of the post-event information. This is because after taking an initial test, the details in the post-event information that relate to the misinformation and consistent questions on the initial test capture attention, and are therefore better encoded. Thus, when taking the final test, the answers to misinformation and consistent questions come to mind more quickly. Because response fluency and confidence are negatively correlated (Kelley & Lindsay, 1993; Koriat, 1993; Koriat, Ma'ayan, & Nussinson, 2006) it was expected that participants would be more confident in their answers to misinformation and consistent items. However, when participants were warned, it was expected that they would no longer respond based on retrieval fluency, and thus the increased confidence that is observed when participants are not warned would not be observed.

The findings of the current study showed that, overall, participants who took a repeated test rated their confidence in their answers as being higher than those who took a single test. The analyses of the effect of warning on each item type separately provides evidence for the retrieval-fluency hypothesis. It was found that confidence in both misinformation and consistent items were lower for participants who were warned than participants who were not warned, but there was no difference in confidence for control items among participants who were warned or not warned. This indicates that misinformation and consistent details in the post-event information captured participants' attention, thus being more easily retrieved, which led to them being more confident in their answers to misinformation and consistent, but not control, questions. These results were expected, as after being warned, participants engage in a more effortful recollection process rather than responding based on retrieval fluency. From the current results it can be theorised that, when warned, participants had to engage in source discrimination, effectively deciding whether the answer they have in mind was part of the original event, or part of the post-event information they read, that they now know may be inaccurate. Thus, participants rated themselves as being more confident in their answers to control questions, as there was no source conflict. Further, participants rated themselves as more confident for consistent than misinformation items, which may suggest that they have a subconscious awareness of a conflict between the accurate details in the video event and the inaccurate details in the post-event information. Further evidence for the retrieval fluency hypothesis lies in the effect of the test condition on confidence ratings. Repeated testing led to higher confidence ratings than single testing, but only for participants who were not warned. This shows that participants who take an initial test, and therefore are open to retrieval-fluency based responding, respond based on how quickly answers come to mind when not warned and therefore report high confidence in their answers, but change their basis of responding to a more effortful recollection when warned. Thus, for participants in the warning condition, the increased confidence they have in their answers due to retrieval-fluency is decreased when they are warned that the post-event information they received may not be accurate.

Results from an analysis of participants' confidence in their responses on misleading trials lends further evidence to the retrieval-fluency hypothesis. When participants provide a response to a question primed with misinformation, we would expect them to respond based on retrieval fluency, and we would expect misinformation answers to come to mind more quickly than correct answers as these details have been better encoded. However, retrieval-fluency based responding would only effect confidence after an initial test, as the test highlights the important details that then 'pop put' while reading the post-event information. Thus, we would expect participants to be more confident when giving misinformation responses, after an initial test. With a warning, however, we would expect participants to revert back to a more careful, effortful recollection, and thus the increased confidence in misinformation responses would no longer be seen. Results from the current study were consistent with predictions. Participants who did not receive a warning had higher confidence when giving a misinformation response than participants who did receive a warning. When broken down into participants who took an initial test and participants who did not take an initial test, a retrieval fluency explanation is consistent with how participants rate their confidence in their answers. For participants who did not take an initial test, there is no difference in confidence for misleading responses when warned and not warned. However, for participants who took an initial test, those who receive a warning rate themselves as being significantly less confident in their answers than participants who do not receive a warning. The results of the current study provide evidence for a retrieval fluency-based explanation, indicating that participants who took an initial test then paid more attention to details in the post-event information that were presented in the test, which led to those details being better encoded. When participants are not warned about the potential inaccuracy of the post-event information, they response based on retrieval fluency when taking the final test. The answers to questions that are based on details that were presented in a misleading way in the post-event information come to mind more quickly for participants who took an initial test, making them more likely to provide a high confidence rating for their misleading responses.

The findings of the current study are in line with previous research on participants confidence ratings associated with the RES effect. Thomas et al. (2010) carried out a similar study and found that repeated testing reduced confidence for misinformation items, but only after participants were warned about the potential inaccuracy of the post-event information. Further, participants were more confident when giving the misinformation response to misleading questions after repeated testing, but only when they are not warned. However, in the Thomas et al (2010) study there was a stronger effect of testing condition. As discussed above, this may be because there is a longer delay in the Thomas et al. (2010) study, and delay has been shown to results in a stronger RES effect (Chan & Langley, 2011).

Response latency findings. According to the retrieval fluency hypothesis, we would expect response latencies for misleading and consistent responses to be faster after an initial test, compared with control items. Further, when warned, we would expect response latencies to be slower, particularly after taking an initial test. The results found that warning participants did not make a difference to their response latencies, which does not support the hypothesis. Also, participants responded faster on control trials than on misinformation and consistent trials, whether they took an initial test or not. According to the retrieval fluency hypothesis, participants should respond faster to misinformation and consistent trials after they have been tested, because the answers are better encoded and thus come to mind more quickly. Although the expected results were found in the study by Thomas et al. (2010), clearly the current study does not replicate these results. As the findings from the recognition accuracy data and the confidence data both supports the retrieval fluency hypothesis, we cannot conclude that the participants in the current study did not respond based on retrieval fluency. Rather, it is possibly due to a lack of power, as discussed above. However, in the Thomas et al. (2010) study the group sizes were comparable to those in the current study, therefore the current study should have sufficient power to be able to detect significant effects, if there are any. A more valid explanation for the limited evidence supporting the retrieval fluency hypothesis in the current study lies in a significant limitation in the measurement of response latency in the current study, such that we cannot draw valid conclusions from the response latency data. This limitation will be discussed later on in this chapter.

The findings from the analysis of the response latencies on misleading trials were consistent with the retrieval fluency hypothesis. The retrieval fluency hypothesis predicts that participants who took an initial test would respond faster when giving an answer consistent with the misinformation, compared to when they give the correct answer. This is because the misinformation answer is better encoded and thus comes to mind more quickly. When warned, however, participants would be expected to slow down their responding because being given a warning encourages them to engage in more effortful processing, which takes longer than responding based on retrieval fluency. On misleading trials, participants have two conflicting sources of information to consider, so it would be expected that responses on these trials should be significantly slower after being warned. The findings of the current study show that in the no warning conditions, participants responded faster (although not significantly) when they gave a misinformation-consistent response than when giving a correct response. However, contrary to predictions, when warned, there was no difference in the response times for misinformation or correct responses. In fact, looking at the raw data it does not appear that warning participants about the potential inaccuracy of the post-event information slowed down their responding at all. We cannot conclude that the warning was ineffective, as warning participants affected recognition accuracy and confidence ratings as predicted. Given the fact that Thomas et al. (2010) found that warning participants *did* have an effect on their response latencies, the lack of an effect in the current study may have been due to lack of power, as discussed above. Finally, as with the response latencies on all trials, response latencies for misleading trials were faster for participants who took an initial test, regardless of whether they gave a correct or misinformation-consistent response. This is consistent with the retrieval fluency account, as repeated testing increases the ease with which participants retrieve the post-event information, as mentioned above. However, as mentioned above, and expanded on below, there are serious limitations with the measurement of the response latency data, so the increase in response latencies with testing does not necessarily infer support for the retrieval fluency hypothesis.

Limitations to the current study and recommendations for future research

An important limitation of the current study was the measurement-related ceiling effect. Participants' average hit rates on the control and consistent questions on the final test were relatively high, with a large number of participants getting every question correct. This resulted in a ceiling effect, with the data being strongly negatively skewed. Because the data for the misinformation questions did not exhibit a ceiling effect, and thus were not skewed, it was not possible to apply an appropriate transformation to the data. As such, the average hit rate data for control and consistent items were not normally distributed, and therefore the conclusions drawn from the average hit rate data may not be accurate (for a more detailed discussion, see Chapter Four). Because a ceiling effect reduces the data variability (Vogt, 2005), it is difficult to detect a significant difference in the average hit rate between groups (Cramer & Howitt, 2004). Thus, future research would need to ensure a ceiling effect in recognition accuracy data was unlikely to occur, either by increasing the difficulty of the questions, or by significantly increasing the number of questions on the test to increase inter-individual variability.

Another significant limitation of the current experiment relates to the response latency measurements. The current study measured response latency as a relatively direct measure of retrieval fluency (Benjamin, Bjork, & Schwartz, 1998). It was thought that repeated testing participants would respond faster to misinformation and consistent items compared with single testing participants, and this would indicate they were responding based on retrieval fluency. However, it is likely that repeated testing participants would respond faster than single testing participants on the final test based on the fact they have already read the questions on the initial test, and thus would be able to read the questions on the final test more quickly, or may not need to read them at all. Thus, we cannot make valid conclusions about retrieval fluency-based responding based on comparisons between repeated testing and single testing participants. Further, it is likely that this is a limitation also in the similar study by Thomas et al. (2010). Thus, while they concluded that the shorter response latencies for those in repeated testing conditions, compared to single testing conditions, provided support for the retrieval-fluency hypothesis, it could in fact have been due to the explanation made above.

Although warning participants about the potential inaccuracy of information may be successful in reducing their susceptibility to misinformation, warnings are likely to be more successful in the laboratory than in real life. In the current study, as in Thomas et al.'s (2010) research, the warning was given after participants had been misled, and before the final test. In the experimental setting this is possible because it is known when and what misleading information is given, but in real-life situations eyewitnesses may be subject to misinformation several times over the course of the investigation, and from many different sources. Thus, in real-life situations it would be difficult to provide a targeted warning to eyewitnesses, so it is unclear how applicable the findings that warning can reduce the RES effect would be in real-life criminal investigations. While so far warning participants is the only known way to reduce the RES effect, given that the use of warnings in real-life situations is not practical, future research needs to focus on finding different, more realistic ways to reduce the RES effect in real-life situations.

Given that the current research failed to find strong evidence for an RES effect, and previous research has shown that the RES effect increases with delay (Chan & Langley, 2011; Chan & LaPaglia, 2011), further research should be carried out to determine how important the delay is between the initial test and the final test in increasing eyewitness' susceptibility to misinformation. In real life situations, there are several different scenarios that are possible with regard to an eyewitness being exposed to misinformation and then retrieving their memory for the initial event. They may retrieve their memory immediately after the event, either by telling their version of events to fellow witnesses, responding emergency personnel, or family and friends (or just rehearsing the event in their heads – I think this is actually a really common and probably much overlooked type of retrieval and is especially likely to occur in stressful and traumatic situations – probably is a fairly haphazard and piecemeal fashion). They may be exposed to misinformation at this point, or maybe much later down the track, possibly by reading or listening to media reports (we could even expose ourselves to 'misinformation' but trying to fill in gaps or imaginatively reconstructing the scene drawing on previous experience and logic). Further, their memory for the event is likely to be tested many times, from immediately after the event to when they are called to give evidence in court possibly months later. As

Chan and LaPaglia (2011) found that multiple tests increase eyewitnesses susceptibility to misinformation compared to a single initial test, and Chan and Langley (2011) found that delay increased eyewitnesses susceptibility to misinformation, further research should be carried out varying the delay between initial and final tests, and the number or tests, to break down the current understanding of the RES effect and find when it is most and least likely to occur.

To the author’s knowledge, there is no existing research that has examined whether the RES effect is found in children. As children are more susceptible to misinformation than adults, particularly young children (Bruck & Ceci, 1999, Ceci & Bruck, 1993), it would be useful to find out how susceptible they are after taking an initial test. Further, while warning participants can eliminate the RES effect in adults (Thomas et al., 2010), it is unclear whether this would have the same effect in children.

Summary and Conclusions

The present study aimed to explore a retrieval fluency explanation for the RES effect and its attenuation with a warning. The study was similar to the Thomas et al. (2010) study, but aimed to explore whether the RES effect would be found using a shorter video event and a shorter delay between the initial test and the presentation of the post-event information.

The current study found limited evidence for the RES effect. Results from the average hit rate data, while in the expected direction, did not reach statistical significance. Results indicated that confidence in both misinformation and consistent items were lower in participants who were warned, compared to participants who were not warned, but there was no difference in control items, suggesting that participants respond based on retrieval fluency. Further, repeated testing led to higher confidence ratings than single testing, but only for participants who were not warned. Results from misinformation trials indicated that for participants in repeated testing conditions, those who received a warning were less confident in their answers to misinformation questions than those who did not receive a warning. Response latency results provide limited evidence for a retrieval fluency explanation for RES, however these response latency findings may not be valid as there is a limitation to the response latency data.

Possible reasons for the lack of statistically significant findings were discussed. The current study used a shorter video event and a shorter delay between the initial test and the presentation of misinformation, and this may be the reason for the lack of an RES effect. That is, the RES effect may be much weaker in situations with a brief event and a short delay before the presentation of misinformation, due to there being less opportunity for forgetting to set in.

References

Ackerman, R., & Koriat, A. (2011). Response latency as a predictor of the accuracy of children's reports. *Journal of Experimental Psychology: Applied, 17*(4), 406-417. doi: 10.1037/a0025129

Akehurst, L., Milne, R., & Kohnken, G. (2003). The effects of children's age and delay on recall in a cognitive or structured interview. *Psychology, Crime & Law, 9,* 97 –107. doi: 10.1080/1068316021000057686

Allen, P., & Bennett, K. (2012). *SPSS statistics: A practical guide version 20.0.* South Melbourne, Australia: Cengage Learning.

Allen, B. P., & Lindsay, D. S. (1998). Amalgamations of memories: Intrusion of information from one event into reports of another. *Applied Cognitive Psychology, 12,* 277-285. doi: 10.1002/SICI)1099-0720(199806)12:3<277::AID-ACP517>3.0.CO;2-B

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review, 5,* 1-21. doi: 10.3758/BF03209454

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*(2)*,* 97-105. doi: 10.1037/h0047507

Belli, R. F., Lindsay, D. S., Gales, M. S., & McCarthy, T. T. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. *Memory & Cognition, 22,* 40-54. doi: 10.3758/BF03202760

Benjamin, A. A., Bjork, R. A., & Schwarz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55-68. doi: 10.1.1.118.1497

Berman, D. E., & Dudai, Y. (2001, March 23). Memory extinction, learning anew, and learning the new: Dissociations in the molecular machinery of learning in cortex. *Science, 291,* 2417-2419. doi: 10.1126/science.1058165

Blank, H. (1998). Memory states and memory tasks: An integrative framework for eyewitness memory and suggestibility. *Memory, 6,* 481-529. doi: 10.1080/096582198388166

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition, 3*(2), 77-88. doi: 10.1037/t22813-000

Brehm, J. W. (1966). *A theory of psychological reactance.* New York, NY: Academic Press.

Brown, A. S., Brown, C. M., Mosbacher, J. L., & Dryden, W. E. (2006). Proactive and retroactive effects of negative suggestion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1234-1243. doi: 10.1037/0278-7393.32.6.1234

Bruck, M., & Ceci, S.J. (1999). The suggestibility of children's memory. *Annual Review of Psychology, 50,* 419–439. doi: 10.1146/annurev.psych.50.1.419

Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403–439. doi: 10.1037/0033-2909.113.3.403

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition, 29,* 1120-1129. doi: 10.3758/BF03206381

Chan, J. C. K., & Langley, M. M. (2011). Paradoxical effects of testing: Retrieval Enhances both Accurate Recall and Suggestibility in Eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 248-255. doi: 10.1037/a0021204

Chan, J. C. K., & LaPaglia, J. A. (2011). The dark side of testing memory: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied, 17*(4), 418-432. doi: 10.1037/a0025147

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,* 6-73. doi: 10.1111/j.1467-9280.2008.02245.x

Christiaansen, R. E., & Ochalek, K. (1983). Editing misleading information from memory: Evidence for the coexistence of original and post-event information. *Memory & Cognition, 11*(5), 467-475. doi: 10.3758/BF03196983

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684. doi:10.1016/s0022-5371(72)80001-x

Cramer, D., & Howitt, D. (2004). *The sage dictionary of statistics: A practical resource for students in the social sciences* (3rd ed.). London, UK: Sage Publications.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken Identification: The eyewitness, psychology and the law.* Cambridge, England: Cambridge University Press.

David, D., & Loftus, E. F. (2005). Age and functioning in the legal system: Perception memory and judgment in victims, witnesses and jurors. In I. Noy & W. Karwowski (Eds.), *Handbook of forensic human factors and ergonomics* (pp. 1-53). London, England: Taylor and Francis.

Dehon, H., & Bredart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but younger adults are more successful at source monitoring. *Psychology and Aging, 19*(1), 191-197. doi: 10.1037/0882-7974.19.1.191

Dodd, D. H., & Bradshaw, J. T. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior, 19,* 695-704. doi: 10.1016/S0022-5371(80)90379-5

Eakin, D. K., Schreiber, T. A., and Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 813–825. doi: 10.1037/0278-7393.29.5.813

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition, 33*(5), 770-782. doi: 10.3758/BF03193073

Fisher, R. P., Geiselman, R. E., & Amador, M. (1989). Field test of the Cognitive Interview: Enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology, 74,* 722 –727. doi: 10.1037/0021- 9010.74.5.722

Frost, P., Ingraham, M., & Wilson, B. (2002). Why misinformation is more likely to be recognised over time: A source monitoring account. *Memory, 10,* 179-185. doi: 10.1080/09658210143000317

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*, 298–307. doi: 10.1007/s10979-008-9146-8.

Gabbert, F., Memon, A., & Allen, K. (2003). Memory conformity: Can eyewitnesses influence each other's memory for an event? *Applied Cognitive Psychology, 17,* 533-543. doi: 10.1002/acp.885

Geiselman, R. E., Fisher, R. P., Cohen, G., Holland, H., & Surtes, L. (1986). Eyewitness responses to leading and misleading questions under the Cognitive Interview. *Journal of Police Science and Administration, 14,* 31 –39. doi: 10.1002/acp.2950

Gieselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *American Journal of Psychology, 99*(3), 385-401. doi: 10.2307/1422492

Goodman, G. S., & Melinder, A. (2007). Child witness research and forensic interviews of young children: A review. *Legal and Criminological Psychology, 12,* 1–19. doi: 10.1348/135532506X156620

Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information*. Journal of Verbal Learning and Verbal Behavior, 21,* 207-219. doi: 10.1016/S0022-5371(82)90571-0

Henkel, L. A., & Franklin, N. (1998). Reality monitoring of physically similar and conceptually related objects. *Memory & Cognition, 26,* 659-673. doi: 10.3758/BF03211386

Higham, P. A. (1998). Believing details known to have been suggested. *British Journal of Psychology, 89,* 265–283. doi: 10.1111/j.2044-8295.1998.tb02684.x

Highhouse, S., & Bottrill, K. V. (1995). The influence of social (mis)information on memory for behavior in an employment interview. *Organizational Behavior and Human Decision Processes, 62,* 220-229. doi: 10.1006/obhd.1995.1045

Hoffmann-Biencourt, A., Lockl, K., Schneider, W., Ackerman, R., & Koriat, A. (2010). Self-paced study time as a cue for recall predictions across school age. *British Journal of Developmental Psychology, 28,* 767–784. doi: 10.1348/026151009X479042

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562–567. doi: 1 0.1016/S0022-5371(71)80029-4

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory, 14,* 47-53. doi: 10.1101/lm.365707

Izawa, C. (1967). Function of test trials in paired-associate learning*. Journal of Experimental Psychology, 75*(2)*,* 194–209. doi: 10.1037/h0024971

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118*(2), 126-135. doi: 10.1037/0096-3445.118.2.126

Judge, M. E., & Quartermain, D. (1982). Characteristics of retrograde amnesia following reactivation of memory in mice. *Physiology of Experimental Psychology: General, 28,* 585-590. doi: 10.1016/0031-9384(82)90034-8

Karpel, M. E., Hoyer, W. J., & Toglia, M. P. (2001). Accuracy and qualities of real and suggested memories: Nonspecific age differences. *Journals of Gerontology: Psychological Sciences, 56B*(2), 103–110. doi: 10.1093/geronb/56.2.P103

Kelley, C. M., & Jacoby, L. L. (1990). The construction of subjective experience: Memory attribution. *Mind & Language, 5,* 49-68. doi: 10.1111/j.1468-0017.1990.tb00152.x

Kelley, C. M., Jacoby, L. L., & Hollingshead, A. (1989). Direct versus indirect tests of memory for source: Judgements of modality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 1101-1108. doi: 10.1037/0278-7393.15.6.1101

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1-24. doi: 10.1006/jmla.1993.1001

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling-of-knowing. *Psychological Review, 100,* 609–639. doi: 10.1037/0033-295X.100.4.609

Koriat, A., & Ackerman, R. (2010a). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science, 13,* 441–453. doi: 10.1111/j.1467-7687.2009.00907.x

Koriat, A., & Ackerman, R. (2010b). Metacognition and mindreading: Judgments of learning for self and other during self-paced study. *Consciousness and Cognition, 19,* 251–264. doi: 10.1016/j.concog.2009 .12.010

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009a). The easily learned, easily-remembered heuristic in children. *Cognitive Development, 24,* 169–182. doi: 10.1016/j.cogdev.2009.01.001

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009b). The memorizing effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology, 102,* 265–279. doi: 10.1016/j.jecp.2008.10.005

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483–502). New York, NY: Guilford Press.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in meta-cognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*, 36–69. doi: 10.1037/0096-3445.135.1.36

LaPaglia, J. A., Wilford, M. M., Rivard, J. R., Chan, J. C. K., Fisher, R. P. (2013). Misleading suggestions can alter later memory reports even following a cognitive interview. *Applied Cognitive Psychology, 28*(1), 1-9. doi: 10.1002/acp.2950

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1077-1083. doi: 10.1037/0278-7393.16.6.1077

Lindsay, D. S., Allen, B. P., Chan, J. C. K., & Dahl, L. C. (2004). Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another event. *Journal of Memory and Language, 50,* 96-111. doi: 10/1016/jjml.2003.08.007

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12,* 361-366. doi: 10.1101/lm.94705

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of memory. *Journal of Experimental Psychology: General, 118*, 100–104. Retrieved from http://www.apa.org/pubs/journals/xge/

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

Loh, W. D. (1981). Psycholegal research: Past and present. *Michigan Law Review, 79,* 659-707. Retrieved from http://www.michiganlawreview.org/

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 371–385. doi: 10.1037/0278-7393.11.2.371

Meade, M. L., & Roediger, H. L., III (2002). Explorations in the social contagion of memory. *Memory & Cognition, 30,* 995-1009. doi: 10.3758/BF03194318

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533. doi: 10.1016/S0022-5371(77)80016-9

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482–488. doi: 10.1037/h0045106

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature, 406,* 722-726. doi:10.1038/35021052

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgements in the allocation of study during multitrial learning. *Psychological Science, 5,* 205-213. doi:10.1111/j.1467-9280.1994.tb00502.x

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125-173). New York, NY: Academic Press

Pallant, J. (2013). *SPSS survival manual* (3rd ed.). New South Wales, AU: Allen & Unwin.

Paterson, H. M., & Kemp, R. I. (2006). Co-witnesses talk: A survey of eyewitness discussion. *Psychology, Crime & Law, 12,* 181-191. doi: 10.1080/10683160512331316334

Przbyslawski, J., Roullet, P., & Sara, S. J. (1999). Attenuation of emotional and nonemotional memories after their reactivation: Role of β adrenergic receptors. *The Journal of Neuroscience, 19,* 6623-6628. Retrieved from http://www.jneurosci.org/

Roediger, H. L., III. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45,* 1043–1056. doi: 10.1037/0003-066X.45.9.1043

Roediger, H. L., III, Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels-of-processing framework. *Memory, 10,* 319–332. doi: 10.1080/09658210224000144

Roediger, H. L., III, & Geraci, L. (2007). Aging and the misinformation effect: A neuropsychological analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 321-334. doi: 10.1037/0278-7393.33.2.321

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255. doi: 10.2307/40064526

Schacter, D., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences, 1,* 229-236. doi: 10.1016/S1364-6613(97)01068-1

Shaw, J. S. III. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied, 2*(2), 126-146. doi: 10.1037/1076-898X.2.2.126

Smith, V. L., & Ellsworth, P. C. (1987). The social psychology of eyewitness accuracy: Misleading questions and communicator expertise. *Journal of Applied Psychology, 72*(2), 292-300. doi: 10.1037/0021-9010.72.2.294

Stevens, J. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York, NY: Routledge.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulated against the build-up of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392-1399. doi: 10.1037/a0013082

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington, DC: National Institute of Justice.

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,* 149-157. doi: 10.1016/j.jml.2010.04.004

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4*(3), 210–221. doi: 10.1037/0278-7393.4.3.210

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*(2), 175–184. doi: 10.1016/S0022-5371(67)80092-6

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behaviour, 13*(2), 181-193. doi: 10.1016/S0022-5371(74)80043-5

Tversky, A., & Kahneman, D. (1972). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232. doi: 10.1016/0010-0285(73)90033-9

Underwood, B. J., & Freund, J. S. (1968). Effect of temporal separation of two tasks on proactive inhibition. *Journal of Experimental Psychology, 78,* 50-54. doi: 10.1037/h0026157

Underwood, J., & Pezdek, K. (1998). Memory suggestibility as an example of the sleeper effect. *Psychonomic Bulletin & Review, 5*(3), 449-453. doi: 10.3758/BF03208820

Vogt, P. (2005). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences* (3rd ed.). London, UK: Sage.

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425,* 616-620. doi: 10.1038/nature01930

Watson, J., McDermott, K., & Balota, D. (2004). Attempting to avoid false memories in the Deese/Roediger-McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32*(1), 135-141. doi: 10.3758/BF03195826

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54,* 277-295. doi: 10.1146/annurev.psych.54.101601.145028

Wheeler, M. A., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*(6), 571–580. doi: 10.1080/09658210244000414

Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3*(4), 240–245. doi: 10.1111/j.1467-9280.1992.tb00036.x

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. A. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29,* 716-732. doi: 10.1016/0749-596x(90)90045-2

Wikipedia. (n.d.). *44 minutes: The North Hollywood shootout.* Retrieved from http://en.wikipedia.org/wiki/44_Minutes:_The_North_Hollywood_Shoot-Out

Wright, D. B. (1993). Misinformation and warnings in eyewitness testimony: A new testing procedure to differentiate explanations. *Memory, 1*(2), 153-166. doi: 10.1080/09658219308258229

Zaragoza, M. S., Belli, R. F., & Payment, K. E. (2007). Misinformation effects and the suggestibility of eyewitness memory. In M. Garry & H. Hayne (Eds.), *Do justice and let the sky fall: Elizabeth F. Loftus and her contributions to science, law and academic freedom* (pp. 35-63). Mahwah, MJ: Lawrence Erlbaum Associates.

Zaragoza, M. S., & Koshmider, J. W., III (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 246-255. doi: 10.1037/0278-7393.15.2.246

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 934-945. doi: 10.1037/0278-7393.20.4.934

List of Appendices

Appendix A: Participant Information Sheet

Eyewitness Memory Study

INFORMATION SHEET

Who is doing this project?

Hello, my name is Amanda Naus and I am currently working on my Master of Science degree in psychology at Massey University. The current research is required for the completion of this degree. The research is being conducted under the supervision of Dr. Stephen Hill. If you have any questions or concerns regarding the research, please feel free to contact either myself or Stephen Hill in the following ways.

Amanda Naus

Phone 0273124549

Email amandajlast@gmail.com

Dr. Stephen Hill

Phone 06 3569099 ext. 7566

Email S.R.Hill@massey.ac.nz

Mail School of Psychology
Massey University
Private Bag 11-222

What is the project about?

This project is interested in how people remember a video event after their memory for the event has been tested. We are not only interested in what you remember accurately, but also the kind of mistakes you may make, as these can tell us about how normal, everyday memory works.

This is an invitation for you to participate in the research. It is assumed that by continuing with this experiment you are giving your consent to participate.

Participant Identification and Recruitment

Approximately 120 participants aged between 18 and 45 are being asked to voluntarily participate in this research. This number is required because participants will be assigned to four experimental groups, and 30 people in each group is necessary to be able to detect patterns and effects in the data.

Participants are being invited to participate in several ways, including through word of mouth, undergraduate Massey classes, and through social media. The only requirements are that you are aged between 18 and 45, are proficient in the English language, have reasonable ability to read, write, and hear (including the use of aids), and can use a computer.

There should be no discomfort or risks to participants as a result of participation. However, participants will be fully debriefed and if any problems arise these can be discussed with either myself or Stephen Hill.

Project Procedures

The project involves watching a short video clip and then carrying out some simple tasks. You will then watch another short video and read a summary of the video. You will then be asked several simple questions about yourself, and then you will take a memory test. Once these tasks have been completed, you will be debriefed.

The time involved for the entire process, including reading information sheets, signing consent forms, and the final debrief, is about 25-30 minutes.

What will happen with the information?

Your responses will be recorded and collated with those of other participants'. The collated results will then be used to evaluate the study questions discussed earlier in this form. All information given will be kept confidential, and individual results will not be personally identifiable in the research findings.

If you would like to be provided with a summary of the overall findings, you may provide your email address or postal address at the end of the experiment.

Participant's Rights

You are under no obligation to accept this invitation. If you decide to participate, you have the right to:

- decline to answer any particular question;
- withdraw from the study;
- ask any questions about the study at any time during participation;
- provide information on the understanding that your name will not be used unless you give permission to the researcher;
- be given access to a summary of the project findings when it is concluded.

For student participants, neither grades nor academic relationships with the School of Psychology or members of staff will be affected by your refusal or agreement to participate.

Ethical approval information

This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named above are responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you wish to raise with someone other than the researcher(s), please contact Professor John O'Neill, Director, Research Ethics, telephone 06 350 5249, email humanethics@massey.ac.nz.

Appendix B: Multi-choice Questions

1. What type of vehicle do the bank robbers drive?
    a. Van
    b. SUV
    c. Station wagon
    d. Sedan

2. What was the name of the bank employee who had the second set of keys?
    a. Hayden
    b. Trevor
    c. Randy
    d. Scott

3. What was the time shown on the bank robber's watch?
    a. 9.01
    b. 9.17
    c. 10.23
    d. 10.30

4. When did the bank robbers first start firing their weapons?
    a. Before entering the bank
    b. Upon entering the bank
    c. Once the employees were down on the floor
    d. After exiting the bank

5. What new station logo was emblazoned on the helicopter?
    a. ABC
    b. Fox 11
    c. CNN
    d. ESPN

6. Where did the bank robbers send the employees?
    a. Bathroom
    b. Staffroom
    c. Vault
    d. Outside

7. What was the name of the bank?
    a. Capital One Bank
    b. Wells Fargo
    c. HSBC
    d. Bank of America

8. There was a lady out walking who came across the bank robbers. What colour was her tracksuit?
    a. Blue
    b. White
    c. Pink
    d. Red

9. Who was the lady walking with?
    a. A woman
    b. A man
    c. A dog
    d. She was walking by herself

10. There is a small business across the car park from the bank that bears the brunt of the bank robbers' bullets. What is the business?
    a. Florist
    b. Food kiosk
    c. Locksmith
    d. Fruit shop

11. What colour was the bank robbers' vehicle?
    a. White
    b. Red
    c. Blue
    d. Silver

12. What was the name of the police department that attended the shootout?
    a. Los Angeles Police Department
    b. New York Police Department
    c. New Orleans Police Department
    d. San Francisco Police Department

13. Where was the first policeman shot?
    a. Chest
    b. Upper leg
    c. Lower leg
    d. Foot

14. What are the two police officer's talking about while patrolling the area, before they notice the bank robbers?
    a. Their weekend
    b. Their kids
    c. An SUV
    d. What they had for breakfast

15. During the shootout, the bullets make a whole in the wall of a neighbouring business. What was this business?
    a. Doctor’s surgery
    b. Dentist’s surgery
    c. Optometrist
    d. Massage therapist

Appendix C: Histograms and Q-Q Normality Plots for Average Hit Rates

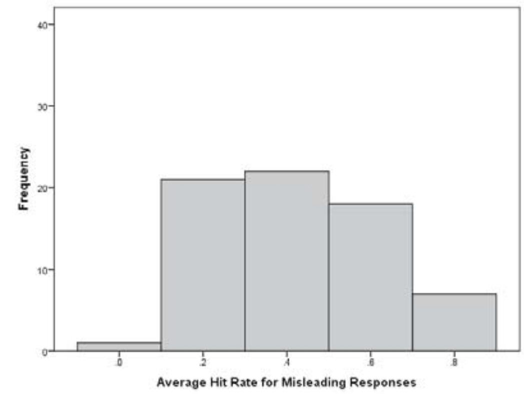


Figure C1: Histogram of average hit rate for misleading responses on the final test

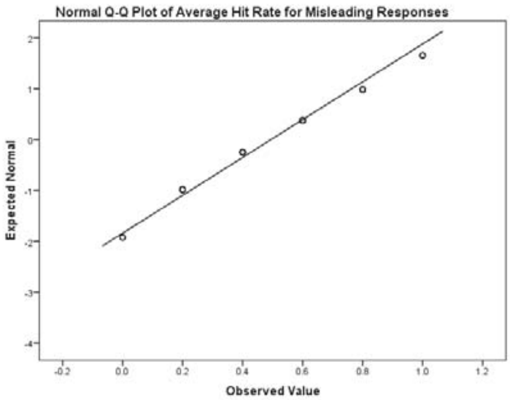


Figure C2: Normal Q-Q plot of average hit rate for misleading responses on the final test

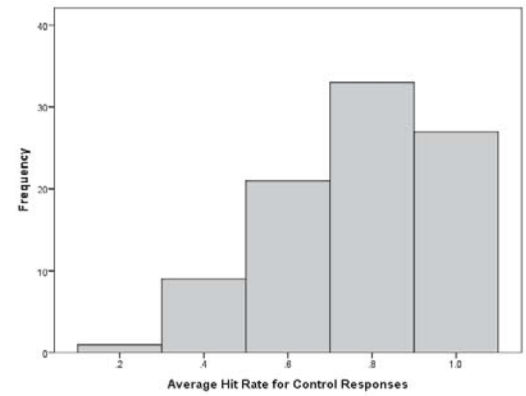


Figure C3. Histogram of average hit rate for control responses on the final test

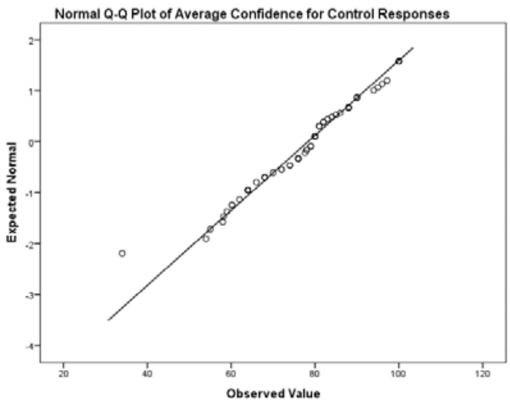


Figure C4. Normal Q-Q plot of average hit rate for control responses on the final test

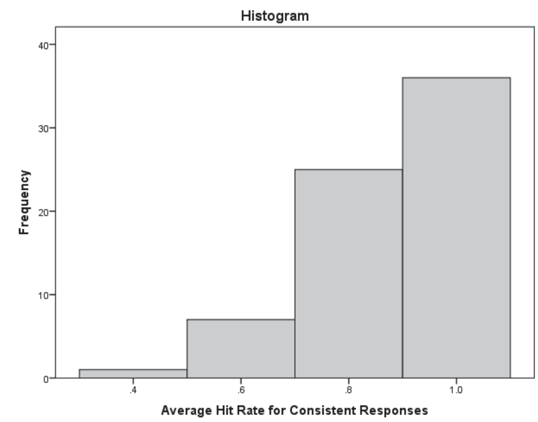


Figure C5. Histogram of average hit rate for consistent responses on the final test

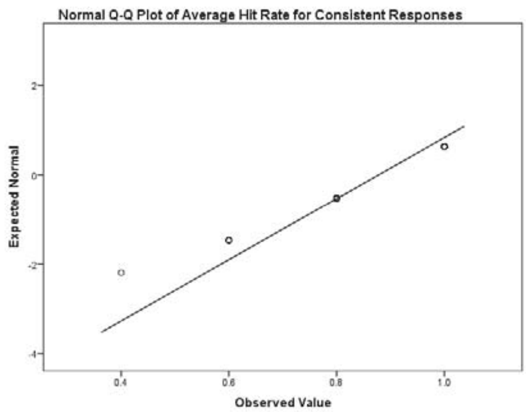


Figure C6. Normal Q-Q plot of average hit rate for consistent responses on the final test

Appendix D: Histograms and Q-Q Normality Plots for False Recognition Rates on Misleading Trials

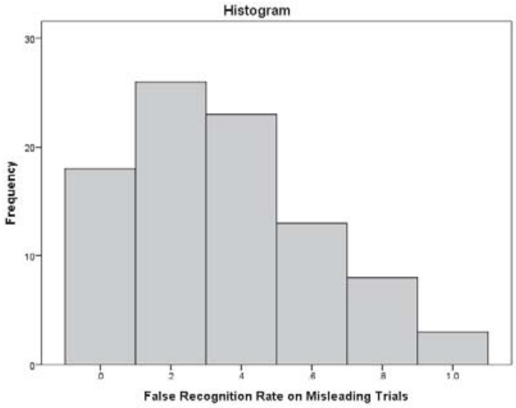


Figure D1: Histogram of average false recognition rates on misleading trials

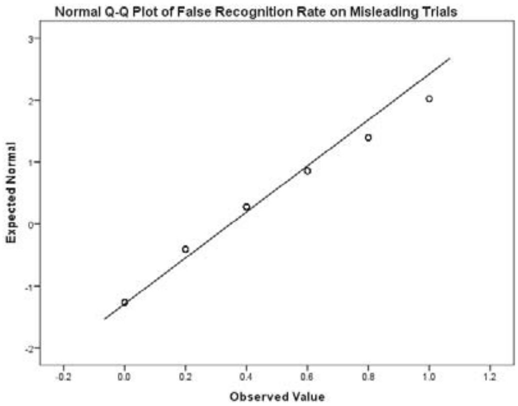
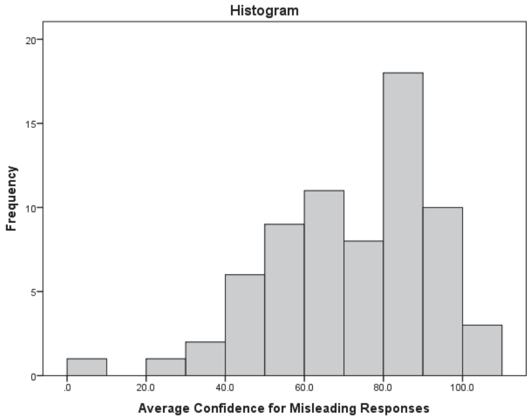


Figure D2. Normal Q-Q plot of average false recognition rates on misleading trials

Appendix E: Histograms and Q-Q Normality Plots for Average Confidence Ratings



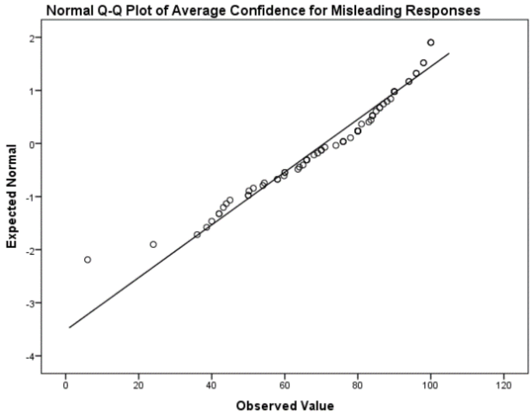
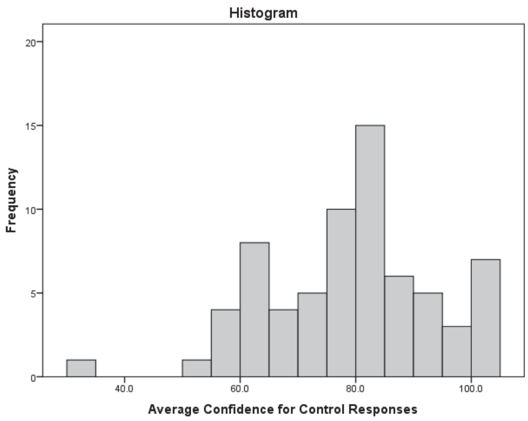


Figure E1. Histogram of average confidence for misleading responses on the final test

Figure E2. Normal Q-Q plot of average confidence for misleading responses on the final test



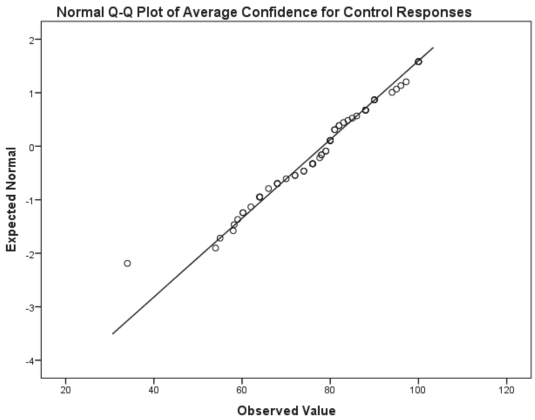
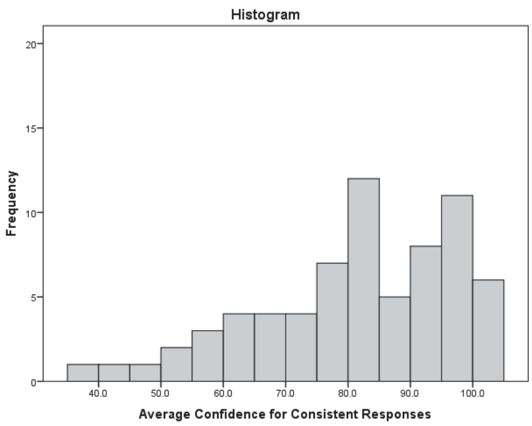


Figure E3. Histogram of average confidence for control responses on the final test

Figure E4. Normal Q-Q plot of average confidence for control responses on the final test



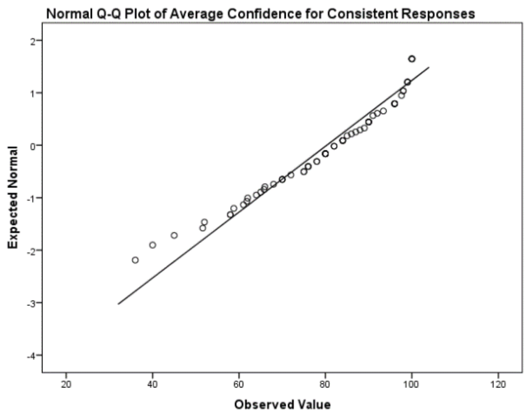


Figure E5. Histogram of average confidence for consistent responses on the final test

Figure E6. Normal Q-Q plot of average confidence for consistent responses on the final test

Appendix F: Histograms and Q-Q Normality Plots for Average Confidence Ratings on Misleading Trials

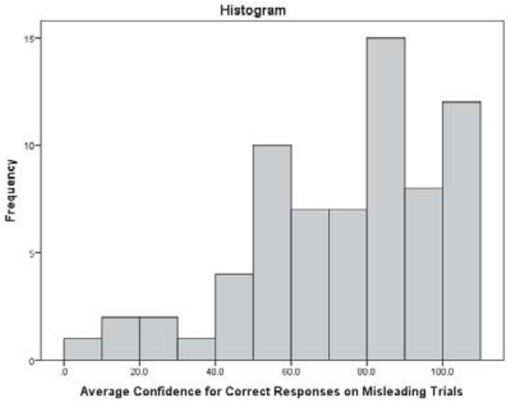


Figure F1. Histogram of average confidence for misinformation responses on misleading trials

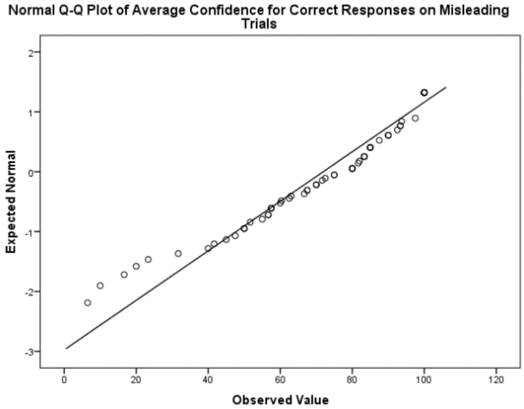


Figure F2. Normal Q-Q plot of average confidence for misinformation responses on misleading trials

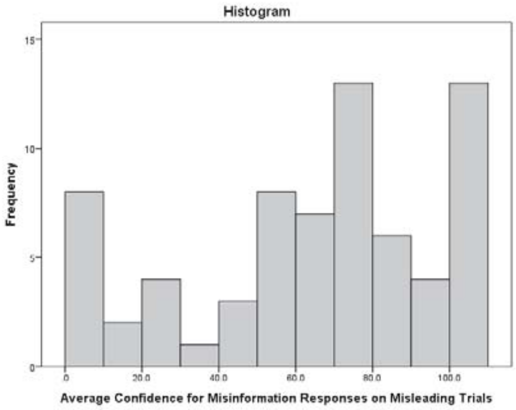


Figure F3. Histogram of average confidence for misinformation responses on misleading trials

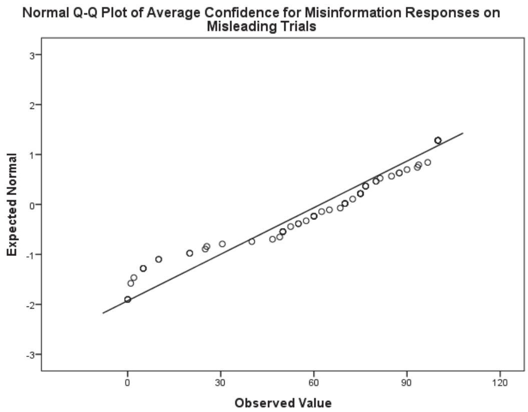


Figure F4. Normal Q-Q plot of average confidence for misinformation responses on misleading trials

Appendix G: Boxplots of Response Latencies

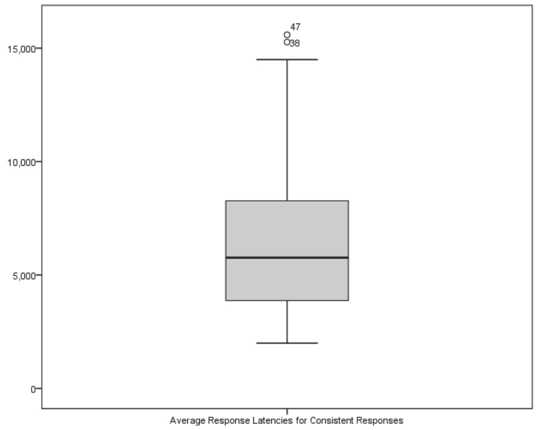


Figure G1. Boxplot of average response latencies for consistent responses on the final test

Figure G2. Boxplot of average response latencies for control responses on the final test

Appendix H: Histograms and Q-Q Normality Plots for Average Response Latencies

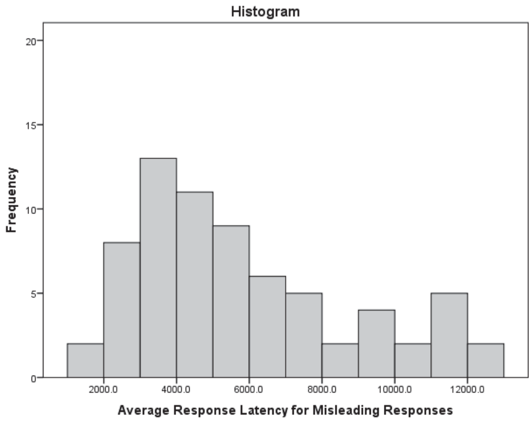


Figure H1. Histogram of average response latencies for misleading responses on the final test

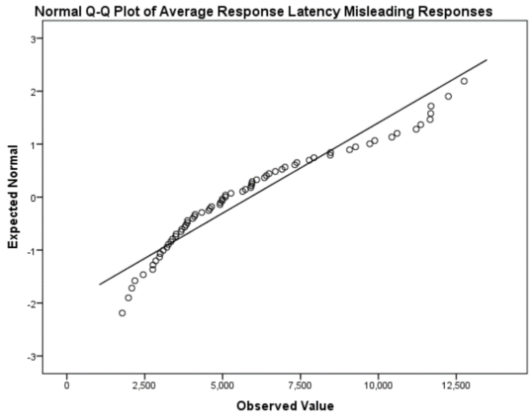


Figure H2. Normal Q-Q plot of average response latencies for misleading responses on the final test

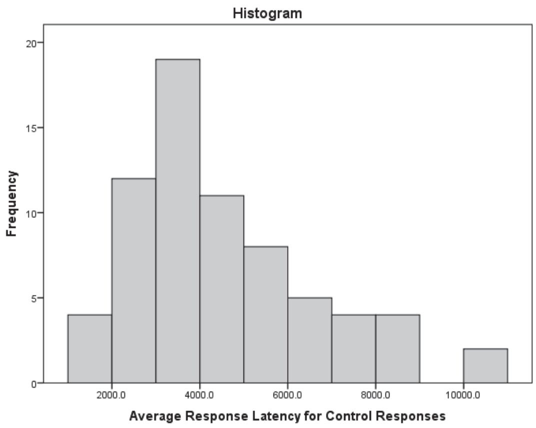


Figure H3. Histogram of average response latencies for control responses on the final test

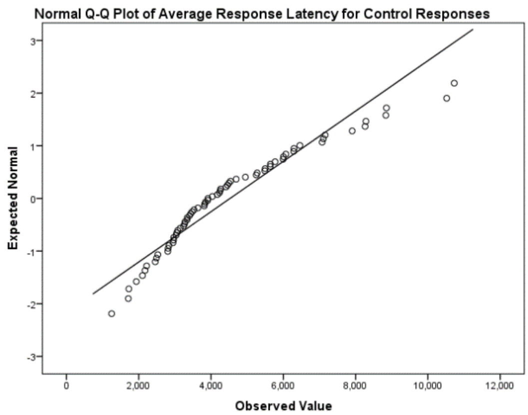


Figure H4. Normal Q-Q plot of average response latencies for control responses on the final test

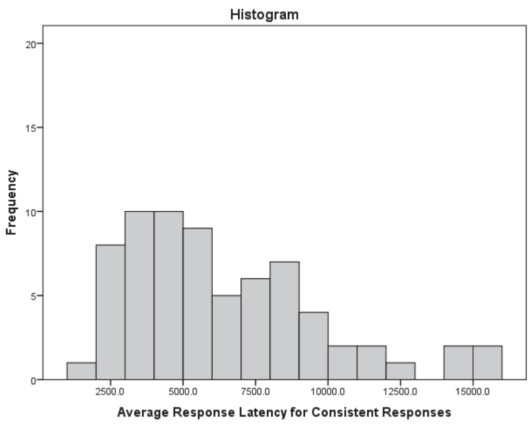


Figure H5. Histogram of average response latencies for consistent responses on the final test

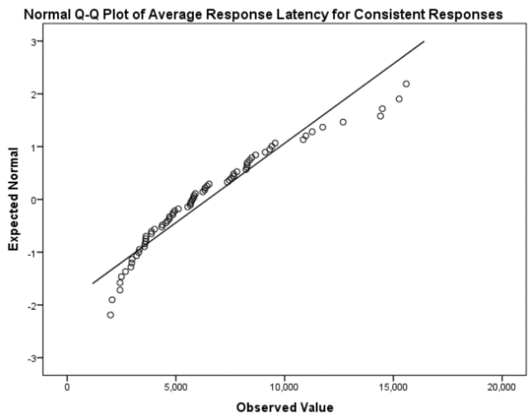
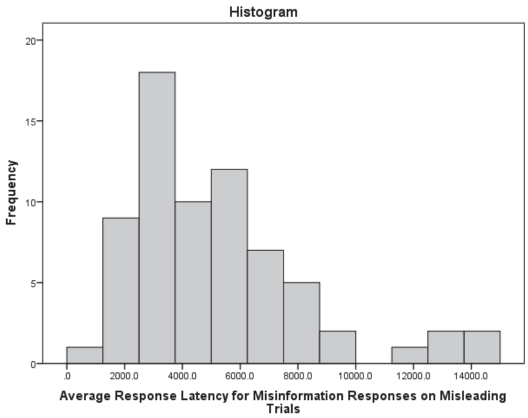


Figure H6. Normal Q-Q plot of average response latencies for consistent responses on the final test

Appendix I: Histograms and Q-Q Normality Plots for Response Latencies on Misleading Trials



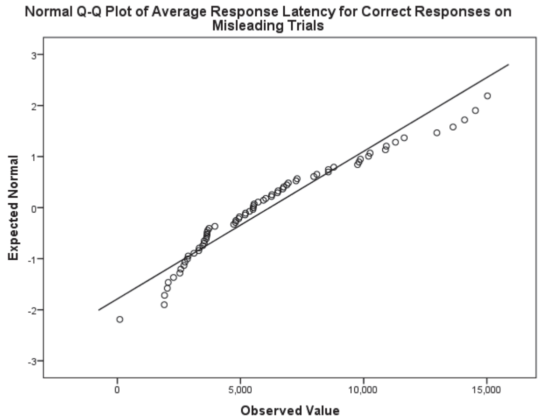
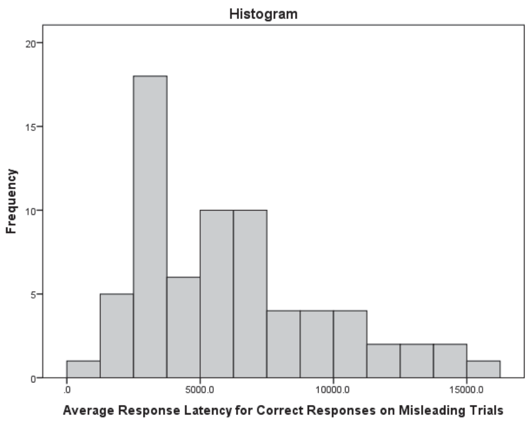


Figure I1. Histogram of average response latencies for misinformation responses on misleading trials

Figure I2. Normal Q-Q plot of average response latencies for misinformation responses on misleading trials



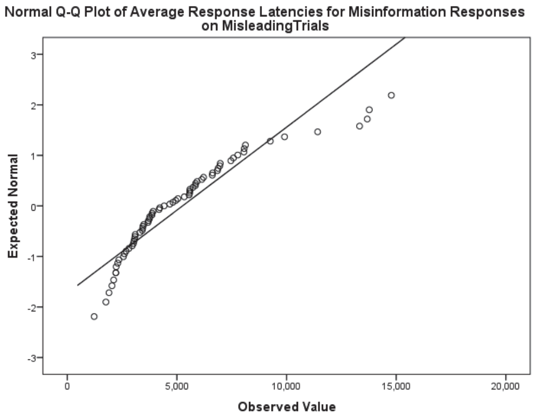


Figure I3. Histogram of average response latencies for correct responses on misleading trials

Figure I4. Normal Q-Q plot of average response latencies for correct responses on misleading trials