Retrieval and Attention:

Factors that Potentiate Learning and Enhance Eyewitness Suggestibility

A dissertation by

Leamarie T. Gordon

In partial fulfillment of the requirements for the degree of

The Doctor of Philosophy

in

Psychology

TUFTS UNIVERSITY

May 2015

Adviser:

Ayanna K. Thomas

Abstract

Research has demonstrated that taking a test prior to receiving misleading information can lead to greater susceptibility to misinformation (Chan, Thomas, & Bulevich, 2009). The present dissertation aimed to fit these findings within a larger literature on *forward effects of testing* (e. g., Tulving & Watkins, 1974; Wissman, Rawson, & Pyc, 2011), where initial testing enhances encoding and learning of subsequently presented information. The present set of experiments examined attention as a possible mechanism underlying the relationship between testing, new learning, and the likelihood of reporting new details in the misinformation paradigm. The pattern of results obtained across experiments indicate that attention to post-event information may impact enhanced suggestibility, and that testing may be one way to promote attention. In contrast, attention was not as important toward facilitating a deeper learning of new, post-test information. These results suggest that retrieval enhanced suggestibility is not merely an indirect example of test-potentiated learning, or forward effects. Rather, retrieval enhanced suggestibility results from a retrieval fluency bias, which may accumulate via an attention mechanism.

Keywords: retrieval enhanced suggestibility, testing, misinformation effects, attention

Acknowledgments

I first want to thank my advisor, Ayanna Thomas, for her endless support throughout my time at Tufts. Thank you for giving me a chance. My life is forever changed because of it. I would also like to thank my dissertation committee members, John Bulevich, Ari Goldberg, Nate Kornell, and Holly Taylor. This project would have been impossible to finish without your insights, questions, and feedback at each stage. I must also thank my previous advisor Amy Shapiro, who has given me so much guidance. To my fellow CAM Lab members. Thank you for always being willing to listen to me at lab meetings. Your support has been amazing. A special thanks to Meeyeon Lee and Amy Smith for making 202 a place not just to work, but also vent and laugh. Tal August and Helen Suntag, thank you for being awesome research assistants. Knowing I could count on you kept me sane. Finally, I would like to thank my family, especially my daughter Kaya Nicole. You are my inspiration and motivation each and every day. I hope I have made you proud. To my sisters, Jessica and Jackie, my brother, Paul, and my mother Arlene and my Nana. I could not have even thought about graduate school without your love and support. Thank you for being there for me and Kaya.

Funding Sources

This dissertation was supported in part by a grant to Leamarie Gordon from the American Psychological Association. Additional support was received from a Tufts Graduate Research Award and the Tufts Psychology Department.

Dedication

This dissertation is dedicated to the memory of my Papa, John Kenneth Gordon, who always believed in me, and to my daughter, Kaya Nicole, whom I will always believe in.

Table of Contents

List of Tables

Introduction

Understanding conditions that lead to memory successes and failures is important, particularly when considering high-stakes situations such as eyewitness memory. Eyewitnesses to crimes are highly valued in the criminal justice system. At early stages of criminal investigations, they are crucial to providing information that leads to arrests. Later, they are called upon to provide sworn testimony, which is often a deciding factor in many criminal trials. Although eyewitnesses serve an important role in the pursuit of justice, it is critical to remember that human memory is far from perfect, and susceptible to distortion. As a striking example of this, since 1989 there have been over 300 convictions overturned due to DNA evidence in the US alone (Innocence Project, 2013). On average, wrongly convicted individuals spend almost 14 years in prison for crimes they did not commit. Overwhelmingly, 75% of the wrongful convictions were due to faulty eyewitness testimony.

Recent research suggests that when eyewitnesses are asked specific questions about a witnessed event, and then are exposed to inaccurate details about the event, they are more likely to incorporate these inaccuracies into later memory reports compared to eyewitnesses who did not immediately answer questions about the event (Chan et al., 2009). This finding, known as retrieval enhanced suggestibility (RES), seemingly contradicts the known benefits that memory retrieval, or testing, typically has on retention and accuracy. Numerous studies in the context of verbal learning and education have demonstrated that testing strengthens memory, and increases the likelihood that the retrieved content will be accessible on later tests (Roediger & Butler, 2011 for review). However in the context of eyewitness memory, retrieval prior to a final memory assessment appears to come with a cost. Recently, Gordon and colleagues (Gordon & Thomas, 2014; Gordon, Thomas, & Bulevich, 2015) argued that RES does indeed fit with a test-enhanced learning framework. Test enhanced learning in most contexts *is* beneficial, but when information learned after an initial retrieval phase conflicts with original memory, such as in the RES paradigm, it manifests as a memory deficit. They propose that immediate testing after a witnessed event initiates a change in processing resources, where more attention is paid to question-relevant information compared to question-irrelevant information (see also Thomas, Bulevich, & Chan, 2010. Attention changes leads to differential processing of the question-relevant material and results in 1) individuals *learning* post-event misinformation better, and 2) the increased likelihood that they will respond with post-event misinformation at final test because these details are highly accessible and more fluently retrieved.

The aim of this dissertation is to carefully evaluate the proposal that RES effects are best categorized as an indirect example of test-potentiated learning, and that testing increases learning and reporting of misinformation via an attention mechanism. While some support exists for role of attention in RES (Gordon & Thomas, 2014; Gordon et al., 2015; Thomas et al., 2010), there are a number of limitations in those studies. First, learning of the post-event narrative has never been directly measured. Second, the conceptualization and operationalization of attention has been limited in scope. Finally, support for the attention allocation hypothesis has only been gathered under conditions where the study of post-event information is self-paced. In these cases, individuals are able to choose how long to attend to post-event information. However, in a number of studies RES has been demonstrated when post-event information is presented in an experimenter-paced format (e. g., Chan et al., 2009; Thomas et al., 2010). The present set of studies addressed these limitations, and tests the test-potentiated learning perspective in two novel and informative ways. In Experiments 1 and 2, attention was directly measured during post-event information processing, and tied to later memory outcomes. In Experiments 3 and 4, participants' ability to differentially process, or review, critical post-event details was disrupted, and post-event learning was measured. In Experiments 2 and 4, learning of the post-event narrative was directly measured on a final memory test.

Eyewitness Suggestibility: The Misinformation Effect

A typical way eyewitness memory is studied in the laboratory is through the misinformation paradigm. In the general paradigm, first developed decades ago by Loftus, Miller, and Burns (1978), individuals watch a video or slideshow depicting a simulated crime (the witnessed event), are then exposed to post-event information that includes misleading details about the event, and after some time take a memory test on details from the witnessed event. The standard finding is that individuals are less accurate on memory tests after exposure to misinformation, and likely to incorporate the misleading details into memory reports (Frenda, Nichols, & Loftus, 2011).

Misinformation Effect Theories. There are two classes of theories to account for the misinformation effect. Some have argued that only one memory trace exists during final test phases. Similar to a neural network approach to catastrophic interference in which existing information is completely forgotten once new information is learned (e. g., McCloskey & Cohen, 1989), the destructive updating hypothesis proposed that stored memories for a witnessed event are completely overwritten by new post-event information (Loftus, Schooler, & Wagenaar, 1985). Misinformation effects occur because the suggested memory is the only one available at the time of test.

Alternatively, a number of theories proposed that two memory traces exist, however individuals are biased at the time of test to respond with misinformation over accurate details. The nature of this bias may vary according to theory. McCloskey and Zaragoza (1985) argued in their strategic effects account that many participants may not have encoded the original event, and while some resort to simple guessing, others may remember the misinformation and choose that at test. Moreover, of those who remember both the original and suggested details, some may be biased toward responding with suggested details on a memory test due to demand characteristics. To pit their hypothesis against the destructive updating account, McCloskey and Zaragoza modified the final memory test so that participants chose between an originally learned item and a novel foil item. They argued that if original memories are completely overwritten by suggested details, then when asked to choose between the original item and a new foil item, participants would be forced to simply guess between the two. Essentially, participants would have no memory for *either* item so final test performance would be at chance. McCloskey and Zaragoza found instead that participants were more likely to choose the true (original) item than the novel foil item on this modified test, suggesting that original memories are not irrevocably altered by misinformation. These results have been replicated in several subsequent studies (e.g., Belli, 1993; Bowman & Zaragoza, 1989; Ceci, Ross, & Toglia, 1987).

Since there is evidence that post-event information doesnot irrevocably alter original event memories, misinformation effects may result from biased responding during final memory tests. Research suggests that biased responding may result from source monitoring errors. Source monitoring errors occur when individuals remember both true and suggested information but confuse the sources of each at test (e. g., Johnson, Hashtroudi, & Lindsay, 1993; Zaragoza & Lane, 1994). In the context of the misinformation paradigm, participants may remember details from the original event and details from the post-event narrative, but incorrectly attribute narrative details to the original event. Source confusions may occur due to unsuccessful encoding processes (e.g., binding of features) that help later discriminate the source of information (e. g., Henkel, Johnson, & de Leonardis, 1998). Alternatively, confusion of sources can occur via a retrieval fluency mechanism. Retrieval fluency refers to the ease with which information is retrieved from memory (Baddeley, 1982; Jacoby & Dallas, 1981). In the context of the misinformation paradigm, participants may simply respond on the final test with whatever information is most accessible in memory in relation to a particular retrieval cue, such as a final test question. In strong support for this theory, Gordon and Shapiro (2012) manipulated accessibility during a final memory test. Participants encoded an event, and then were exposed to misleading information. Prior to presenting participants with a recognition test where they decided between a true and misleading item, some participants were primed with semantic associates of the true item. They found that primed participants were more likely to choose the true item than participants who were not primed. That is, the misinformation effect was reduced by increasing the accessibility of originally learned details. In further support of a fluency bias account, when participants are placed under experimental conditions that encourage a more careful evaluation of source memory during the final test phase, and discourage fluency-based responding, the misinformation effect is diminished (Bekerian & Bowers, 1983; Christiaansen & Ochalek, 1983; Kroll, Ogawa, & Nieters, 1988).

Repeated Retrieval and Misinformation. The generally agreement that misinformation effects result from breakdown of effective source monitoring has theoretical importance. This discovery has directly translated retroactive interference findings from traditional verbal learning studies into memory for complex events. Verbal learning retroactive interference paradigms require individuals to study one list of words, then a second list of words, and later recall words from only List 1 (e. g., Müller and Pilzecker, 1900). Generally, poorer recall of List 1 words is observed when List 2 is interpolated between List 1 study and final test (see Wixted, 2004 for review). That is, List 2 learning interferes with List 1 recall. In the case of the misinformation paradigm, List 2 (e. g., the misleading details) interferes List 1 recall (e. g., the witnessed event). Practically, this research has been influential in the legal arena, by highlighting the malleability of human memory and cautioning against the accuracy of eyewitness memory reports.

While research conducted using the standard misinformation paradigm has proved informative and influential, the paradigm may not best reflect what happens in the real world when an individual witnesses and event and then encounters post-event information. In an authentic eyewitness situation, witnesses are required to retrieve original event details multiple times. They may call 9-1-1, talk to other bystanders, or call friends and family. In each of these scenarios, the eyewitness practices retrieval of event details, prior to exposure to secondhand post-event information, and before official testimony is provided. This initial retrieval is notably absent from the traditional misinformation paradigm generally used to study misinformation susceptibility (e. g., Loftus et al., 1978). Thus, until recently it was largely unknown what effect initial retrieval would have on eyewitness memory accuracy.

Chan et al. (2009) looked at repeated retrieval in the misinformation paradigm. Participants first viewed an event depicting fictitious criminal activity, and then were split into two groups. A *repeated test group* took an immediate cued recall test on details from the event. A *standard misinformation group* completed an unrelated filler activity for an equivalent amount of time. Both groups then listened to a post-event narrative that provided a synopsis of the event. Importantly, the narrative included consistent details (details that were repeated from the event), neutral details (details that did not contradict the event, but were non-specific), and misleading details (details that directly contradicted information from the event). Based on the numerous findings that demonstrate that testing is beneficial to memory (Roediger & Butler, 2011 for review), they predicted that initial retrieval would strengthen eyewitness's memory for the original event, and buffer against the effects of post-event misleading details. Specifically, they expected that the repeated test group would have better final memory test performance than the standard misinformation group. In other words, the standard group would demonstrate a misinformation effect and the repeated test group would not.

Interestingly, these hypotheses were not supported. Instead of finding that immediate retrieval led to better memory for the witnessed event and reduced susceptibility to misinformation, Chan et al. found that the repeated test group demonstrated *larger* misinformation effects than the standard group, a finding they termed *retrieval enhanced suggestibility* (RES; see also Chan & Langley, 2011; Chan &

LaPaglia, 2011; Gordon, Cernasov, Bulevich, & Thomas, 2015; Gordon et al., 2015; Gordon & Thomas, 2014; Thomas et al., 2010). Importantly, the RES effect is generally characterized by two separate findings. The first is a decrease in accuracy on misleading trials on a final memory test. That is, relative to the standard group participants in the repeated test group report fewer correct answers on the final test after exposure to misinformation. The second is increased production of misleading details. When individuals are incorrect on final test trials associated with misinformation, repeated test participants are more likely to report the misleading details compared to the standard test group.

RES Theories. As with standard misinformation effects, there are two classes of theories to explain RES. Reminiscent of the destructive updating hypothesis of standard misinformation effects, Chan and colleagues (Chan et al., 2009; Chan & LaPaglia, 2013; Chan, Wilford, & Hughes, 2012) have argued that RES results from a reconsolidation-like mechanism (c. f., Nader & Hardt, 2009) whereby access to original event information is disrupted by testing. According to the account, immediate retrieval of the original event activates specific event details, placing them into labile states where they are highly susceptible to interference from contradictory post-event details. Original details are no longer accessible and the misleading details are the only ones available in memory during later retrieval attempts.

Chan and LaPaglia (2013) tested the reconsolidation hypothesis by implementing a source-free recognition test as the final memory test. After viewing a fictitious crime event, a repeated test group took a cued recall test assessing video memory, while a standard misinformation group completed an unrelated filler task. Both groups were then presented with a narrative synopsis of the video, and then took a final source-free recognition test. On this final test, participants were presented with details from either the original event or the narrative, and were instructed to respond “old” if they remembered the information from *either* source, and to respond “new” otherwise. Chan and LaPaglia (2013) argued that this source-free testing procedure produced a strong estimate of memory accessibility, because participants were able to indicate that information was previously presented regardless of whether they remembered it occurring in the original event or post-event narrative. Under these testing procedures, repeated test participants were less likely to recognize details from the original event as compared to participants who did not take the initial test. That is, when repeated test participants were asked about an original event detail, they responded that these items were “new”. There was no between-group difference in recognizing narrative details as “old”.

Although the reconsolidation account explains both the accuracy and production components of RES, immediate testing does not always render the original event information inaccessible (e. g., Thomas et al., 2010; Gordon & Thomas 2014; Thomas, Chen, Gordon, & Tenbrink, 2015). In fact, when certain final test procedures are implemented, a standard benefit of testing on memory is revealed on misleading trials. For example Thomas et al. (2010) found that when participants were warned about the veracity of the narrative before final testing, those who had taken an initial test before presentation of the narrative were *more* likely to retrieve original event information on the final test than participants who had not taken an initial test. Further, Gordon and Thomas (2014) demonstrated that when participants were encouraged to report two answers in association with a given final test question, participants who had taken an initial test were more likely to report details from the narrative *and* video compared to participants who had not taken an immediate test. These results provide strong evidence against a reconsolidation/inaccessibility account. It is also important to note that reduced final test accuracy after testing is not consistently found within RES studies (Chan & Langley, 2011; Chan et al., 2012; Wilford, Chan & Tuhn, 2014). Increased production of misleading details after testing, on the other hand, is *always* found. Theory development, then, should primarily focus on the production effect in RES.

To account for these findings, Gordon and colleagues (Gordon & Thomas, 2014; Gordon et al., 2015) have argued that RES effects do accompany two memory traces, and that individuals have representations intact for both post-event information and original information. Similar to a bias account of misinformation effects, RES results because retrieval magnifies this bias. They propose that after testing, participants *learn* post-event information more efficiently than individuals who were not initially tested. Thus, while testing may strengthen original memory, the test-potentiated new learning (e. g., misinformation) is more accessible during later memory tests. Newly learned misleading details are likely to be produced on later tests, because the ease with which test-potentiated narrative learning comes to mind biases responding even if original event details are still available in memory (cf., Baddeley, 1982; Jacoby & Dallas, 1981). Several findings support this theory. First, Thomas et al. (2010) demonstrated that participants who took an immediate test demonstrated quicker decision times when choosing misleading details on a final four-alternative forced choice test compared to participants in a standard misinformation group. Second, Thomas et al. (2010) found that higher confidence judgments accompanied final test answers on misleading and consistent trials for participants who had taken an immediate test as compared to participants who had not been immediately tested. Research has consistently demonstrated that the ease with which information comes to mind serves as an indicator for confidence (Kelley & Lindsay, 1993; Nelson & Narens, 1990). Third, when participants were encouraged to discount retrieval fluency, those that had taken an immediate test demonstrated better memory for the original event details as compared to those in a standard misinformation group (Thomas et al., 2010; Gordon & Thomas, 2014). Thus, in contrast to the reconsolidation account, RES may be considered an instance of test potentiated learning. Enhanced learning may occur because after testing individuals differentially attend to and encode misleading narrative details. In turn, this process increases the fluency with which narrative details are subsequently retrieved and biases final test responding.

Testing, Learning, and Memory

The present dissertation examines RES as an example of test-potentiated learning. The perspective taken is grounded in broader theories focused on the impact of retrieval on memory and learning outside of the misinformation paradigm.

Backward Effects of Testing. One of the most well-documented and long-standing findings in cognitive psychology is that repeatedly retrieving learned information leads to better retention of the information than other methods, such as simple restudying (Carrier & Pashler, 1992; Gates, 1917; Roediger & Karpicke, 2006). That is, testing reinforces the strength of the memory for the retrieved material.

An early account of this standard testing effect proposed that testing simply re-exposes individuals to learned material, which then strengthens memory representations. A number of studies effectively discounted this idea after finding that individuals who took repeated tests on studied material had better long-term retention than individuals given multiple opportunities to study the same material (Larsen, Butler, & Roediger, 2009; McDaniel, Anderson, Derbish, & Morrisette, 2007; Wheeler, Ewers, & Buonanno, 2003).

Subsequent theories focused on specifying the role of retrieval in transforming memory. One way that retrieval may strengthen memory is through the elaboration of the memory trace and the creation of additional retrieval routes, which increases the likelihood of successful future retrieval (but see Karpicke & Smith, 2012). According to the *elaborative retrieval hypothesis* (ERH; Carpenter, 2009, 2011; Carpenter & DeLosh, 2006) presentation of a retrieval cue activates information in memory that is related to the cue during the search for the target. The additional activated information is encoded along with the cue-target information, providing additional retrieval routes through which to access the target on later memory tests. Pyc and Rawson's (2010) *mediator effectiveness hypothesis* more specifically defined 'additional information'. They proposed that tests enhance the link between a cue and target via mediating information, or a word or concept that links a cue to a target. For example, a cue (*mother*) and target (*child*) are semantically linked by the mediator (*father*). When participants learned such cue-target pairs through testing, Carpenter (2011) found that they were more likely to false alarm to mediators during a later recognition test, and more likely to recall a target word when given the mediator as the cue.

Forward Effects of Testing. In addition to the basic *backward effect* of testing, in which testing strengthens existing memory representations (see Roediger & Butler, 2011 for review), recent reviews have highlighted other beneficial effects of testing on memory. For example, testing can promote the transfer of learning to novel contexts (Carpenter, 2012). Specifically, initial tests have been shown to improve performance on later tests of different formats (Kang, McDermott, & Roediger; 2007), and have promoted memory for related content that was never tested (Chan, McDermott, & Roediger, 2006).

Research has also unveiled *forward effects* of testing, or enhanced learning of *new* information presented after retrieval. In verbal learning studies, learning of new information after initial study is augmented when participants are first tested on initially studied material (Arnold & McDermott, 2013; Izawa, 1966; Szpunar, McDermott, & Roediger, 2008; Tulving & Watkins, 1974). For example, in an early verbal learning study, Tulving and Watkins (1974) demonstrated greater learning of A–C in an A–B, A–C paired-associate learning paradigm if a test of A–B preceded A–C learning, as compared to when no test intervened between the learning phases. More recently, Wissman et al. (2011) extended this work to educationally relevant prose material and demonstrated that test questions interpolated between text passages facilitated learning of material presented in subsequent passages that was related to the tested material. Finally, Kornell and colleagues (Hays, Kornell, & Bjork, 2013; Kornell, Hays, & Bjork, 2009; Richland, Kornell, & Kao, 2009) demonstrated that tests can facilitate subsequent study episodes of relevant information, even when initial retrieval has failed.

Limited theories explain how testing potentiates learning of new information, and have primarily been evaluated in verbal learning studies. Two factors appear to play a role: retrieval and encoding. Retrieval explanations typically assume that testing between the study of words lists or word pairs enhances list differentiation and reduces interference between lists at test (Szpunar et al., 2008; Bäuml and Kliegl, 2013). Proactive interference occurs when information learned in early study phases interferes with one's ability to learn and recall information from later study phases (Postman & Keppel, 1977). One way list differentiation may be promoted via testing is through the enhancement of source memory, or the ability to correctly attribute learned information to the context in which it was acquired. When original learning is retrieved, source specifying information is also retrieved and strengthened. At later retrieval attempts, individuals are aided by this source information and better able to discriminate between learning episodes. When proactive interference is successfully overcome in such a way, enhanced recall of subsequently learned information, or proactive *facilitation*, is often observed (Postman, 1962).

Encoding explanations, on the other hand, argue that testing facilitates learning of new material because it improves the *encoding* of the material. On an unconscious level, encoding may be facilitated via activation of related information during initial testing (c. f., Carpenter, 2011; Chan et al., 2006; Grimaldi & Karpicke, 2012; Hays et al., 2013). Similar to the elaborative retrieval explanation of backward effects, memory retrieval activates a target and related information, and this activation in turn facilitates the incorporation of new information into memory. This can be thought of as a form of long-term semantic priming that can then be boosted by conscious processes, such as attention

(c. f., Balota, Black, & Cheney, 1992; Becker, Moscovitch, Behrmann, & Joordens, 1997; Chan et al., 2006). On a more conscious level, testing may serve to ‘reset’ the encoding process, such that post-test information is encoded as effectively as pre-test information (Pastötter, Schicker, Niedernhuber, & Bäuml, 2011). Alternatively, testing may result in a change in participants’ encoding strategy, enhancing encoding for post-test information relative to pre-test information (e. g., Wissman et al., 2011).

RES as a Forward Effect

As mentioned, RES accompanies two distinct behaviors. The first is a reduction in final test accuracy, which is a not always found. The second is an increased in production of misleading details, which is always found. It is important to understand the distinction between accuracy and production because, on the surface, RES contradicts what is known about beneficial backward effects of testing on memory. Gordon and colleagues (Gordon & Thomas, 2014; Gordon et al., 2015) have argued RES does indeed fit with a test-enhanced learning framework. While the backward effect of testing in the misinformation paradigm is not immediately apparent, they argue that the forward effect of testing is. That is, the production effect may be the result of test-potentiated learning of *new*, misleading details. The misleading details are then erroneously reported on later memory tests because participants are biased to report them via a retrieval fluency mechanism.

While there is some support for test-potentiated learning in the misinformation paradigm (Gordon & Thomas, 2014), and evidence for a fluency bias at final test (Thomas et al., 2010), the relationship between testing and retrieval fluency is not well specified. How does retrieval, or testing, influence the accumulation of retrieval fluency? To address this question, Gordon and colleagues recently proposed the *attention* *allocation hypothesis*, or the idea that immediate testing after a witnessed event initiates a shift in processing resources, where more attention is paid to cue-relevant information compared to cue-irrelevant information (Gordon & Thomas, 2014; Gordon et al., 2015). This proposal parallels an encoding account of forward testing effects in general learning studies (c. f., Carpenter, 2011; Chan et al., 2006; Grimaldi & Karpicke, 2012; Hays et al., 2013; Pastötter et al., 2011; Wissman et al., 2011).

In support of the attention allocation hypothesis, Gordon and Thomas (2014) found participants spent more time reading sentences in a post-event narrative that contained misleading details about a witnessed event than sentences containing neutral details. This difference was greater in a group that took an initial test on the witnessed event compared to a group that did not take an initial test. Gordon et al. (2015) extended these findings by demonstrating that more time was spent reading sentences that included details directly relevant to initial test questions (either consistent with or contradictory to the encoding event) compared to neutral sentences. A contingency analysis based on participants' performance on initial test questions revealed that when participants were *correct* on initial test questions, they spent more time reading misleading details that contradicted their responses. Finally, Gordon et al. yoked narrative processing to final test output. Recall of misleading details on the final cued recall test were associated with more time processing misleading sentences compared to trials where repeated-test participants reported some other wrong answer. This difference was not present in the standard group.

Taken together, these studies indicate that information presented after testing in the misinformation paradigm is better attended to, which leads to be improved learning as demonstrated by increased misinformation production effects on subsequent memory tests. However, the evidence in support of the attention allocation hypothesis has been provided *only* through the collection of reading times during post-event narrative processing. In a number of RES studies, including the original study (e. g., Chan & Langley, 2011; Chan et al., 2009; Thomas et al., 2010), participants *listened* to the post-event narrative. In these experiments, reading time data cannot account for demonstrations of the RES effect because the narrative is experimenter-paced. When the narrative is presented in a written format, and participants' reading times of individual sentences are measured, it is at the discretion of individual participants to decide how much time to spend processing each sentence. Alternatively, when the narrative is presented in an experimenter-paced aural manner, participants cannot choose to spend more time listening to, or processing, some sentences relative to others. Yet, enhanced learning of misinformation is still observed. One of the primary goals of this dissertation is to test the hypothesis that even in the context of experimenter-paced narrative presentations, participants are able to spend additional time thinking about, or reviewing, misleading details after they are presented. Thus, this dissertation broadens and better specifies the current and rather limited conceptualization of the attention allocation hypothesis.

Attention. Previous research provides some support that attention allocation to post-event information changes in response to immediate testing in the RES paradigm. These changes are associated with enhanced learning of details in the post-event narrative. However, when the focus is broadened to include other instances of test-potentiated learning (e.g., those found in paired-associate learning and text passage learning), the general consensus is that neither backward nor forward testing effects can be explained by an attention mechanism. For example, Shapiro and Gordon (2012) compared a condition in which students were given in-class questions targeting lecture material, to a condition in which targeted lecture material was presented in a red flashing font and students were explicitly told that the material was important to an upcoming exam. They found that students' exam performance was better for material that was initially tested relative to material presented in a way designed to grab attention. Richland et al. (2009) gave participants pretest questions and then presented a passage for study that could be used to answer those questions. Participants who were given pre-test questions demonstrated better learning of the passage on a final test compared to participants who were given the passage with important details in highlighting type-face.

There are important distinctions to be made when it comes to attention, particularly when considering its relationship to memory. Attention is not always a bottom-up process. It is not always 'grabbed' by exogenous factors such as physical changes to a stimulus, such as in Shapiro and Gordon (2012) and Richland et al. (2009). Attention can also be 'grabbed' in a top down manner and consciously controlled by an individual, a common distinction made in the attention literature (e.g., McCormick, 1997; Stolz, 1996; Theeuwes, 1991; Yantis & Jonides, 1984). This type of endogenously driven attention has a particularly important bidirectional relationship with memory. Both endogenous *and* exogenous attention impact what will be remembered through selection of information and modulation of encoding. In other words, attention impacts memory because information that is differentially attended to is encoded more efficiently, supporting retention. On the other hand, memory impacts endogenous attention only, by optimizing the selection of information through past experience (Chun & Turk-Browne, 2007). Memory impacts attention because individuals can use prior knowledge and experience to ‘select’ what should be preferentially attended to.

As such, it may be inappropriate to discount the role of attention in test-enhanced learning based on comparisons made between two vastly different attention direction mechanisms (e. g., Shapiro & Gordon, 2012; Richland et al., 2009). Memory retrieval, or testing, is likely to provoke *endogenous* attention. A retrieval cue, or test question, is a vehicle for prior knowledge and experience. When an individual is presented with a test question and attempts to answer it, more time may be spent reviewing the information relevant to the question. Moreover, memory for the test question and one’s response to the question may guide future learning episodes, and draw attention to information relevant to the question/response. This may be particularly relevant in the RES paradigm. Not only might participants’ memory for test questions and responses guide subsequent information processing, when participants encounter narrative details that contradict what they responded with, this too may capture attention (c. f., Gordon et al., 2015). Attended narrative details are differentially processed, better learned, and more fluently accessed on later retrieval attempts.

In addition to attention being multi-faceted, it is generally understood that attention is limited in capacity. It is impossible to devote full attention to all possible environmental and internal stimuli at once. According to the limited capacity model of attention (e. g., Kahneman, 1973; Lynch & Srull, 1982) the total attentional capacity allocated to process all activities can be divided into two parts: capacity devoted to a primary task and spare capacity. Generally, there is some spare capacity, or available cognitive resources such as attention, to complete a primary task while successfully completing a secondary task. When the demands of one task increase, however, performance on the latter task will decrease proportionately to the increased demand of the primary task (Knowles, 1963; Moray, 1967). When one task is cognitively demanding and requires effort and attention, then the distribution of processing resources becomes more inequitable and efficient performance on a simultaneous task decreases even further. The impact of a secondary task on performance of a primary task is dependent on the structure of the secondary task (Wickens, 1980). Wicken's (2008) multidimensional resource allocation model defines resources according to three dimensions: stage of processing (perception/cognition or response), code of processing (spatial or verbal), and modality (visual or auditory). The model predicts that the more dimensions shared by concurrent tasks, the worse the performance. For example, simultaneous performance on a visual–verbal task and an auditory–verbal task (which share the code of processing dimension) should be worse than performance on a visuospatial task and an auditory–verbal task (which differ along said dimension). In other words, some secondary tasks will interfere with primary tasks more than others.

A number of studies over the years have shown that increasing attention to an item or task can interfere with the ability to process other information (Kinchla, 1992; Pashler, 1998). In the context of the attention allocation hypothesis of retrieval enhanced suggestibility, initial testing maybe one way to increase attention to the primary task of processing post-event narrative details. When the mind is occupied with the primary task of thinking about, or reviewing, recently encoded post-event narrative details, it follows that performance on a secondary task, such as detecting a visual probe, will slow.

Moreover, when manipulations that *increase* attention to the primary task are employed, such as initial testing, the response times on the secondary task should be reduced even further (c. f., Reynolds & Anderson, 1982).

The Present Study

The primary goals of the present study were two-fold. First, in these experiments I investigated the *forward effect of testing* explanation of retrieval enhanced suggestibility, and directly tested whether testing facilitates new learning in the misinformation paradigm. Second, I examined an encoding account of the forward effect, and specifically the role attention plays in facilitating encoding. Prior studies suggest that after taking a test on a witnessed event, participants differentially attend to misinformation presented in a written post-event narrative. It is this test-induced change in attention that promotes learning and increased retrieval fluency of post-event narrative details (Gordon & Thomas, 2014; Gordon et al., 2015; Thomas et al., 2010). However, these studies provide only indirect support for some of their claims. The experiments presented here address these issues.

First, studies that have looked at the relationship between attention and test-enhanced learning outside of the RES paradigm have limited the conceptualization of attention. Such studies have minimized the role of attention (e. g., Richland et al., 2009; Shapiro & Gordon, 2012), and have only considered attention as a construct manipulated by exogenous factors. Here, the perspective is taken that testing endogenously directs attention. Moreover, within the RES paradigm, support for the attention allocation hypothesis has only been gathered under conditions where the study of post-event information is self-paced. Experiments 1 and 2 were designed to measure attention when the post-event narrative is presented in an experimenter-paced, aural manner. Finally, the forward effects account is directly tested in Experiments 2 and 4. During the final test phase, participants were required to report what was learned in the narrative, rather than what was witnessed during the original event as in typical RES studies.

In addition to these primary aims, this set of experiments had a number of secondary goals. First, the experiments provided an initial investigation into an alternative explanation of RES. Although presently untested within the context of eyewitness memory, research on the *hypercorrection effect* could be important to understanding final test performance in the RES paradigm. In a typical hypercorrection experiment, participants are given a test, often general knowledge questions, and asked to provide answers to the questions and confidence ratings for each of their responses. Participants are then given feedback on their performance. If they were correct they are informed as such, if they were incorrect the correct answer is provided. After some delay, participants are presented with the same set of questions to answer again. The hypercorrection effect occurs when individuals are more likely to correct high confidence errors made on the first test than errors held with low levels of confidence (Butterfield and Metcalfe, 2001; 2006). This procedure is comparable to the RES paradigm. In the RES paradigm, some participants take an initial test. Although they are not explicitly told they are correct or incorrect on the initial test, the post-event narrative may serve as a form of performance feedback. Given an initial test question is answered with the correct video details, consistent narrative details may serve as confirmatory feedback, while misleading narrative details may serve as corrective feedback. To test the hypercorrection account, in each experiment confidence ratings were collected on the initial memory tests, and tied to later misinformation production.

Second, in Experiment 1, the final test included questions that were only semantically and contextually related to initial test questions. This manipulation provided the opportunity to inform theories of test-enhanced learning in a larger sense. Specifically, if backward effects of testing occur because retrieval cues activate information in memory related to the cue which leads to elaboration of the memory trace (e. g., Carpenter, 2009, 2011; Carpenter & DeLosh, 2006), then it follows that after initial testing memory performance for the 'related' final test question should increase. Finally, in each experiment, after taking the final memory test participants were re-exposed to each test question. They were instructed to indicate whether they remembered different details from the narrative and video that could answer each question. Comparisons of discrepancy recollection between a group who takes an immediate test and a group who does not may provide additional evidence against a reconsolidation account of RES. A reconsolidation account would predict that repeated test participants would be less likely to report a discrepancy on misleading trials because they no longer have access to video details. If repeated test participants do not show a disadvantage on this task, then it indicates that video memories are still accessible in this group.

Experiment 1

Experiment 1 was designed to extend the attention allocation hypothesis of RES to account for situations in which participants cannot choose how much time is spent directly processing post-event information. The post-event narrative was presented in an aural, experimenter-paced manner. I tested the hypothesis that after testing, participants may differentially allocate attention to misleading details presented in the narrative by spending time reviewing, or rehearsing, the details. According to the limited capacity model of attention (e. g., Kahneman, 1973; Lynch & Srull, 1982), reaction times to an unrelated, secondary task during post-event narrative processing reflect the strength of attention given to the narrative, even in a situation where direct processing time cannot be measured. If initial testing serves to increased attention to question-relevant information in the post-event narrative in such a way, then a difference should be observed in secondary task reaction times between participants who are initially tested and participants who are not initially tested, when they must make responses to the task following the presentation of question-relevant information.

In addition, a hypercorrection account of RES was examined through an analysis based on Test 1 confidence ratings. If hypercorrection effects play a role in RES, then greater susceptibility to misinformation should be observed on trials where participants reported answers with high confidence. To explore an elaborative retrieval account of the backward effects of testing, participants' performance on a subset of final test questions that were related to initial test questions, but whose answers could not be learned in the post-event narrative, was examined. If elaborative retrieval underlies testing effects, then individuals who take an initial test should be more accurate on related final test questions than individuals who did not take an initial test. Finally, recollection of discrepancies between video and narrative content were examined. I predicted that repeated test participants would perform no worse than participants who did not take an immediate test on the discrepancy recollection task. This will provide evidence against a reconsolidation account. If participants, after testing, are unable to access video memory, they should not be able to accurately recollect a discrepancy between the video and narrative.

Experiment 1 Method

Participants

A total of 58 undergraduates from Tufts University participated. A total sample size of 60 was chosen based on previous studies with the same factorial design (e. g., Chan et al., 2009; Chan & Langley, 2011; Gordon & Thomas, 2014). Two participants were removed from the study for failure to comply with instructions. Participants were compensated either with course credit or were paid $15.

Materials and Procedure

Group: (Repeated Test, Standard Misinformation) was manipulated between subjects while Item Type: (Consistent, Neutral, Misleading) was manipulated within subjects. All participants first watched a 22-minute video clip of the black and white silent film “Rififi”, from now on referred to as the encoding event. The clip portrayed a group of four men committing a burglary in the middle of the night. No participant had viewed the video before.

Following the video, participants in the repeated test group took an immediate cued recall test on details from the video. Twenty four questions were constructed as initial test stimuli. Test questions were validated for appropriateness in a separate pilot study. Pilot participants watched the video, then answered a number of test question. Questions were chosen as stimuli for the present experiments if between 40 and 70 percent of participants answered them correctly. Test questions were cued recall questions requiring brief one or two word response, and targeted specific details from the video (e. g., *What type of jewelry did the man remove from the drawer of valuables?*). Test questions were presented in the same order as events were depicted in the video. After answering each initial test question, repeated test participants provided a rating for how confident they were in their answer (0-Complete Guess to 100-Completely Confident). Instead of taking the initial test, the standard misinformation group single test group completed a Sudoku puzzle for an equivalent amount of time (6 minutes). All participants then completed a brief demographic questionnaire, and a synonym and antonym vocabulary test (Salthouse, 1993). These tasks were not used in any analysis, but were used to fill a brief retention interval. Following the initial phase, all participants listened to an audio recording of a narrative describing the events from the video. Critically, the narrative contained 24 sentences that introduced consistent, neutral, and misleading information about the video (8 details each), in addition to filler information that was not tested in either the initial or final test phases. Consistent sentences contained details that were accurate regarding the encoding event (e. g., *From a drawer that holds valuables, he removes a ring.*) Neutral information was accurate regarding the video, but non-specific. These details could not be used to accurately answer test questions (e. g., *From a drawer that holds valuables, he removes a piece of jewelry*.) Misleading information always consisted of a detail in the video that had been changed in the narrative (e. g., *From a drawer that holds valuables, he removes a necklace.*) Sentences serving as misleading, neutral, and consistent were counterbalanced across participants.

Participants only listened to one sentence pertaining to each critical detail, and the critical detail (ring/jewelry/necklace) was always the last word in the sentence.

During the narrative presentation, all participants were instructed to view a computer screen. Periodically throughout the narrative, an on-screen cue (a non-verbal symbol: <) appeared alerting participants to make a keyboard response. Reaction times for these responses were recorded. If a cue appeared on the left side of the screen, participants were instructed to press a key of the right-hand side of the keyboard. If a cue appeared on the right side of the screen, participants will be instructed to press a key of the left-hand side of the keyboard. A response was coded as *correct* if the appropriate key was pressed in response to the visual cue. Importantly, cues appeared only *after* the presentation of the 24 sentences that included critical information. The visual cues appeared 1.5 seconds after the presentation of the critical detail in each sentence had completed. Importantly, the following sentence was always a filler sentence that was accurate regarding the video, but contained no critical information tested at any point during the experiment. This ensured that processing of critical information was not inadvertently disrupted during the stimulus-response task.

Following the narrative phase, all participants took a final cued recall memory test. The 24 questions from the initial test phase were presented in the same order as the events in the video. In addition to the 24 questions that were on the initial test, an additional 18 questions were added to the end of final test. These questions were related to the initial test question via the context of the original video. For example, an initial test question asked “What does assailant D use to mark where he'll drill on the safe?” The related question asked “How many spots does Assailant D mark to drill into on the safe?" Participants were instructed to respond only with information they remembered from the original encoding event (video). They were encouraged to answer each question, but were told if they did not have an answer they could withhold a response. Following the final test phase, participants completed a discrepancy detection task. Here, they were re-presented with each final test question, and required to respond (yes/no) whether they remembered different details presented in the video and narrative in association with each question. Participants were then thanked and debriefed.

Experiment 1 Results

Reaction Time to Probe Task

The reaction time analyses were limited to trials on which the response to the probe was correct. Participants were either accurate or inaccurate on a given probe trail. Accurate responses were made when the participants pressed the correct computer key in response to the visual probe. This task did not change as a function of item type. However, it is important to determine that performance was similar across item types. On average, participants responded correctly to the probe when presented after consistent details 89% of the time, after neutral details 88% of the time, and after misleading details 87% of the time. There was no difference in probe accuracy across trial types, $p = .79$.

Reaction times to the probe task were trimmed using the following parameters. Reaction times faster than 200ms, and slower than 4000ms were excluded from analysis. These parameters indicate non-compliance with task instructions. The reaction time data for each item type were checked for normality using the Shapiro-Wilks test. The data were determined non-normal on consistent trials ($W = .88$, $p < .01$), neutral trials, ($W = .91$, $p < .01$), and misleading trials ($W = .92$, $p < .01$). Prior to analysis, the data were transformed to fit a normal distribution using a natural log transformation. Using the log-transformed reaction time values, a 3 (Item Type: Consistent, Neutral, Misleading) x 2 (Test Group: Repeated, Standard) ANOVA was conducted. The analysis revealed a main effect of item type, F (2, 112) = 3.74, p < .05. Participants were slower to respond to the probe after misleading trials compared to consistent trials, t (57) = 2.63, p < .01, d = .34. The comparisons between consistent and neutral (p = .15) and neutral and misleading trials (p = .17) were not significant. Table 1 presents log-transformed mean reaction times.

An analysis was also conducted on the raw data using median reaction times. A 3 (Item Type: Consistent, Neutral, Misleading) x 2 (Test Group: Repeated, Standard) ANOVA was conducted. Neither the main effect of item type (p = .14) or testing group (p = .51) nor the testing group by item type interaction (p = .74) was significant. Table 2 reports median reaction times.

Memory Performance

All follow up comparisons used a Bonferroni correction unless otherwise stated. Accurate recall on both Tests 1 and 2 were calculated by dividing the total number of trials in which participants produced the correct video detail out of the total number of trials for that given item type. This was a binary measure, where participants were either correct or incorrect. During the initial recall test, .54 of participants' responses were accurate and .06 produced misinformation spontaneously.

Accurate Video Recall on Final Test. Table 3 presents the accurate recall probabilities on the final test. A 3 (Item Type: Consistent, Neutral, Misleading) x 2 (Test Group: Repeated, Standard) analysis of variance (ANOVA) examined the effects of item type and testing on final accurate recall. It revealed a main effect of item type, F (2, 112) = 70.96, $p < .001$, $\eta_p^2 = .56$. Participants were most accurate on consistent questions (M = .79) as compared to neutral (M = .54) and misleading question trials (M = .42). Planned comparisons revealed that consistent trials resulted in significantly greater accuracy as compared to control, $t(57) = 10.81$, $p < .01$, $d = 1.49$. In addition, participants were more accurate on control trials as compared to misleading trials, $t(57) = 3.74$, $p < .01$, $d = .49$. Finally, participants were more accurate on consistent compared to misleading trials, $t(57) = 10.07$, $p < .01$, $d = 1.32$. The main effect for testing group ($p = .85$) and the item type by testing group interaction ($p = .33$) were not significant.

Misinformation Production on Final Test. Table 4 reports probabilities of misinformation production on the final test. Misinformation production was calculated by dividing the number of misleading trials in which participants produced the misleading detail out of the total number of misleading trials. A between-participants t-test revealed no difference in misinformation production between the standard misinformation group (M = .41) and the repeated test group (M = .47), $p = .41$.

Final Test Confidence. Separate analyses were conducted on final test confidence ratings for correctly reported video details, and for inaccurate misleading details. A 2 (Testing Group: Standard, Repeated) x 3 (Item Type: Consistent, Neutral, Misleading) ANOVA on confidence on questions answered correctly was not significant, all ps > .19. A t test comparing testing groups' confidence ratings for produced misleading details was not significant, $p = .45$. Mean confidence ratings are reported in Table 5.

Related Test Questions

An independent *t* test assessed whether performance on related final test questions would differ as a function of the presence of initial testing. The difference between the standard group (*M* = .54) and repeated test group (*M* = .55) was not significant.

Discrepancy Recollection

An independent samples *t* test assessed the effect of testing on the ability to detect a discrepancy between details presented in the video and narrative on misleading trials. There was no difference between the standard (*M* = .53) and repeated test (*M* = .50) groups, *p* = .62. Means are reported in Table 6.

Test 1 Confidence and Final Test Performance

The Test 1 confidence analyses were limited to the repeated test group because they were the only participants to take the initial test. These analyses were conducted to explore the role hypercorrection effects may have in test-enhanced learning in the RES paradigm. Hypercorrection occurs when individuals answer a question incorrectly, receive feedback on the question, and are more likely to correct high confidence errors made on a first test than errors held with low levels of confidence (Butterfield and Metcalfe, 2001; 2006). This can be looked at in multiple ways in the present study. Considering hypercorrection from a traditional sense, in the present study when participants are *incorrect* on Test 1, correct details in the narrative may serve as feedback. More relevant to whether a similar type of process contributes to RES are behaviors observed on misleading trials, which can be looked at in two ways. First, when participants are *incorrect* on Test 1, *misleading* details may also serve as Test 1 feedback.

Second, when participants are *correct* on Test 1, misleading details can serve as feedback. Consistent and misleading trials types are presented here separately.

During the initial test, participants were instructed to type in confidence ratings on a scale of 0 to 100. On occasion, participants typed in non-numeric responses, or omitted responses. To quantify this, across participants in Experiment 1, there were 720 test 1 confidence responses possible. Of these, 97% were valid. Participants contributed unequally, and in some cases not at all, to cell means based on testing group, trial type, and final test response type. To increase power and weight participants' contributions to each analysis more appropriately, the data was pooled across participants and an item analysis was conducted. That is, each valid Test 1 confidence rating was yoked to its accompanying Test 1 and 2 accuracy scores, and served as an individual data point.

Consistent Trials. Looking only at consistent trials on which participants were incorrect on Test 1, a *t* test with Test 2 video accuracy (correct, incorrect) as the independent variable was performed on mean Test 1 confidence ratings. The test was not significant. When correct responses were provided on Test 2, the associated Test 1 confidence ratings (n = 57; M = 42.1) were similar to instances in which incorrect responses were provided on Test 2 (n = 46; M = 41.6). In other words, hypercorrection was not observed.

Misleading Trials. More relevant to RES are the analyses on misleading trials. To test whether participants' Test 1 confidence was associated with the likelihood of endorsing misinformation when test 1 responses were incorrect, a *t* test with Test 2 misinformation production (produced, not produced) as the independent variable was performed on mean confidence ratings. The test was not significant. When misleading responses were produced on Test 2, the associated Test 1 confidence ratings (*n* = 64; *M* = 32.9) were similar to instances when misleading details were not provided on Test 2 (*n* = 41; *M* = 38.8).

To test whether participants' test 1 confidence was associated with the likelihood of endorsing misinformation when test 1 responses were *correct*, a *t* test with Test 2 misinformation production (produced, not produced) as the independent variable was performed on mean confidence ratings. The test was significant, t (126) = 2.54, p = .01, d = .45. When correct Test 1 answers were given, those that later turned to misleading answers on Test 2 were associated with lower confidence (*n* = 47; *M* = 69.5) than those that remained correct on Test 2 (*n* = 81; *M* = 82.9).

Experiment 1 Discussion

Experiment 1 sought to extend the attention allocation explanation of RES to account for situations in which misleading post-event information is presented in an experimenter-paced format. No difference was observed between the standard and repeated test group on reaction times to the probe task on any trial type. One interpretation of these data is that participants did not allocate more attentional resources to post-event narrative processing after testing. A similar manipulation was effective at measuring attention to a primary task in Reynolds and Anderson (1982). However, in contrast to the present experiment Reynolds and Anderson presented post-test information visually in blocks of text, and the secondary task was auditory. It is possible that in the present paradigm the manipulation was not sensitive enough to reveal attention allocation differences.

In addition, misinformation effects were demonstrated in both the standard and repeated testing group. Misinformation was demonstrated in the context of video-memory accuracy and narrative detail production. Importantly, RES was not demonstrated. It is important to consider that a between-group difference was absent *not* because the repeated test group did not demonstrate a large production effect, but because the standard group demonstrated a similarly large effect. Two prior RES studies using and audio narrative and final cued recall test found standard test groups report misleading details on an average of 27% of misleading trials (Chan et al., 2009; Thomas et al., 2010). To compare, in the present experiment this same group reported misleading details on 41% of misleading trials. Thus, the standard group demonstrated enhanced suggestibility, in the absence of initial retrieval.

What set the present experiment apart from other RES experiments was the addition of the probe response task during narrative presentation. The intention was to index attention to the primary narrative task, without disrupting actual processing of the narrative. As clearly demonstrated by the production effects in the standard group, the probe task did not disrupt narrative processing. On the contrary, it appears to have enhanced it. While this pattern was unexpected, studies reporting a new attentional phenomenon may be useful here. The attention boost effect (ABE) is the surprising finding that when two continuous tasks are performed concurrently, for example listening to a narrative while responding to a visual probe, a momentary increase in attention to one task (probe) enhances, rather than impairs, performance in another task (narrative encoding). This effect has been demonstrated with both visual and verbal materials (Mulligan, Spataro, & Picklesimer, 2014; Spataro, Mulligan, & Rossi-Arnaud, 2013;

Swallow & Jiang, 2010;). ABE contrasts with the majority of the dual-task performance and attention literature by showing that increasing attention to one task can trigger an attentional process that supplements, rather than impairs, performance on a second task. It is possible that the probe task triggered an attentional process similar to that triggered by testing, leading to similar levels of misinformation encoding and production. Such a process could explain the lack of reaction time differences between the standard and repeated test groups. This attention effect was not additive, however, since repeated test participants also completed the probe task and did not demonstrate narrative learning beyond that of the standard group. This might suggest that the attentional processes elicited by testing and the probe task are of the same endogenous type.

Another major finding regarded the discrepancy recollection measure. Initial testing did not impair participants' ability to recollect that a conflicting detail was presented in the original video. That is, repeated test participants were similarly likely to report discrepancies compared to the standard group. Contrary to what a reconsolidation account of RES would predict, initial testing followed by interfering information did not reduce accessibility to the original memory representation. Rather, it increased accessibility to interfering information.

Finally, Experiment 1 examined the potential contributions of hypercorrection to RES, and elaborative retrieval to backward effects of testing in a broader sense. Hypercorrection effects were not observed in the present study. In fact, the opposite pattern emerged. When participants were correct on the initial test, responses that were later changed to misleading details on the final test were associated with lower Test 1 confidence than responses that were not later changed. Relative to elaborative retrieval, repeated test participants performed no better on final test questions that were not presented on test 1, but were related to Test 1, than participants who did not answer initial test questions. Thus, an elaborative retrieval mechanism did not facilitate retrieval of previously learned material in the present experiment.

Experiment 2

Experiment 1 primarily looked at the role of attention in RES effects. These effects necessarily emerge when participants are required to respond with video memory on a final test. Experiment 2 was designed to specifically look at the role of attention in test enhanced learning by directing participants to respond with post-event narrative memory on the final test. As the final test assessed memory for the narrative, and Test 1 still assessed memory for the video, hypercorrection could not be examined in Experiment 2. Further, elaborative retrieval could not be examined. Related question content was only presented in the context of the video, thus participants were not be able to answer with narrative details on related question trials.

Experiment 2 Method

Participants

A total of 58 undergraduates from Tufts University participated and were compensated either with course credit or $15.

Materials and Procedure

The materials and procedure were identical to Experiment 1 with one change. On the final test, participants were instructed to report what they remembered learning in the narrative only. They were allowed to omit responding.

Experiment 2 Results

Reaction Time to Probe Task

As in Experiment 1, the reaction time analyses were limited to trials on which the response to the probe was correct. On average, participants responded correctly to the probe when presented after consistent details 93% of the time, after neutral details 90% of the time, and after misleading details 87% of the time. There was no difference in probe accuracy across trial types, $p = .41$.

Reaction times to the probe task were trimmed using the same parameters as in Experiment 1. The reaction time data for each item type were checked for normality using the Shapiro-Wilks test. The data were determined non-normal on consistent trials (W =.77, $p < .01$), neutral trials, ($W = .86$, $p < .01$), and misleading trials ($W = .87$, $p < .01$). Prior to analysis, the data were transformed to fit a normal distribution using a natural log transformation. Using the log-transformed reaction time values, a 3 (Item Type: Consistent, Neutral, Misleading) x 2 (Test Group: Repeated, Standard) ANOVA was conducted. No effects were significant, all ps > .18.

An analysis was also conducted on the raw data using median reaction times. A 3 (Item Type: Consistent, Neutral, Misleading) x 2 (Test Group: Repeated, Standard)

ANOVA was conducted. No effects were significant, all *p*s > .14. Table 2 reports median reaction times.

Memory Performance

During the initial recall test, .62 of participants' responses were accurate and .07 produced misinformation spontaneously.

Final Test Narrative Recall. On the final test, participants were instructed to report information they learned in the narrative only. For consistent trials, this was a video detail that was then repeated in the narrative. For misleading trials, this was a new detail presented only in the narrative. Neutral trials did not provide specific details in the narrative, so they are excluded from analysis here. Narrative recall of repeated (consistent) and new (misleading) details were examined separately. It was important to examine consistent and misleading trials separately, because item repetition on consistent trials would confound any trial type comparison on narrative recall. Consistent details were presented in the video and then repeated in the narrative. Misleading details, on the other hand, were 1) never presented in the context of the video and 2) were inconsistent with the video. Table 7 reports proportions of accurate narrative recall across trial types.

Consistent Trials. An independent *t* test compared testing groups on mean recall of consistent narrative details. The repeated test group (M = .87) recalled significantly more details than the standard test group (M = .71), t (56) = 3.09, p < .01, d = .84.

Misleading Trials. An independent *t* test compared testing groups on mean recall of new, misleading narrative details. The repeated test group (M = .71) recalled significantly more details than the standard test group (M = .43), t (56) = 4.18, p < .001, d = 1.08.

Final Test Confidence. An independent *t* test compared testing groups on confidence ratings for correctly recalled consistent narrative details. The repeated test group reported higher confidence on these trials, t (56) = 3.04, $p < .01$, $d = .81$. The same comparison was not significant on correctly recalled misleading narrative details, $p = .31$. Mean final test confidence ratings for Experiments 2 and 4 (final narrative test) are reported in Table 8.

Discrepancy Recollection

An independent samples *t* test revealed a significant difference between the standard ($M = .47$) and repeated test ($M = .66$) groups on ability to recollect a discrepancy between details presented in the video and narrative on misleading trials, $t(56) = 3.30$, $p < .01$, $d = .87$. Means are reported in Table 6.

Experiments 2 Discussion

Experiment 2 investigated the role of attention in test-potentiated learning of post-event narrative information. Several interesting findings emerged. As in the first experiment, there was no difference between the standard and repeated test groups in reaction time to the probe task, indicating that attention processes during presentation of the narrative were similar in each group. On the final test of memory for the narrative, however, group differences emerged. Across item types, the repeated test group reported more correct narrative details than the standard group, demonstrating a forward effect of testing. More important to the relationship between forward effects and RES, initial testing potentiated the learning of new, inconsistent event details. That is, repeated test participants recalled more misleading details from the narrative than standard participants. The pattern of results across Experiments 1 and 2 indicate a disassociation between the cognitive mechanisms that underlie retrieval enhanced suggestibility and forward effects of testing. While the probe task may have boosted attention in the standard test group, leading to similar misinformation effects in each testing group in Experiment 1, that same attention boost did not increase new narrative learning in the standard group in Experiment 2. Standard test participants in Experiment 2 recalled fewer misleading narrative details when specifically directed to do so compared to the repeated test group.

A second between-testing group difference emerged in Experiment 2. Repeated test participants were more likely than standard participants to recollect a discrepancy between what was presented in the video and what was encountered in the narrative on misleading trials. This difference was not significant in Experiment 1. The difference in detection recollection in the repeated test groups across the two experiments may be due in part to a change in the demands of the final test. In both experiments, original event encoding was followed by a *seemingly* accurate retelling of the video. A final test in which instructions highlight the differences between the narrative and video, such as specifically instructing participants to respond with narrative details only, may encourage more careful evaluation of source. A final test that assesses memory for the original video would not necessarily engender the same careful discrimination between the two sources. The standard group also took this more demanding source test in Experiment 2 and still recalled fewer discrepancies relative to the repeated test group. Repeated testing itself can promote source memory (e. g., Jang & Huber, 2008; Pastötter et al., 2011; Szpunar et al., 2008). Thus, the repeated test group in Experiment 2, who benefitted from both initial testing *and* a more demanding source test, would be better equipped to both report misleading narrative details and recollect discrepancies between the video and narrative (c. f., Wahlheim, 2015). The pattern of results across Experiments 1 and 2 will be elaborated upon in the General Discussion.

Experiment 3

Experiments 1 and 2 were designed to measure attention to post-event narrative details. The experiments demonstrated that the probe manipulation may have encouraged endogenous attention in the standard test group in a similar manner as the initial test manipulation. This changed accessibility to and the likelihood of reporting narrative details (Experiment 1), but did not promote the association of learned narrative details and source (Experiment 2). Retrieval practice, or initial testing, may be required for the latter process.

Experiments 3 examined the contribution of attention to suggestibility in an alternative way. If attention is required for narrative learning and RES, then inhibiting participants' ability to preferentially process misleading narrative details should reduce RES. To do so, Experiment 3 strategically manipulated attention to post-event narrative details through a disruption task.

Experiment 3 Method

Participants

The number of between group factors in Experiment 3 differed from prior RES studies. Thus, a sample size estimation was calculated using G*Power Version 3 software. Using moderate parameters (power = .8, effect size f = .25) the analysis estimated a sample size of 124. A total of 124 undergraduates from Tufts University participated and were compensated either with course credit or $15.

Materials and Procedure

The experiment used a 2 (Testing Group: Standard, Repeated) x 2 (Attention: Full, Divided) x 3 (Item type: Consistent, Neutral, Misleading) mixed design. Testing group was manipulated between subjects, while processing and item type served as within subject factors. Participants first watched the Rififi video, and then took an immediate cued recall test on the video (repeated test group) or played Sudoku (standard group). Confidence ratings for each test 1 response were collected in the repeated test group. The same stimuli for the initial and final test phases will be used as in Experiment 1. Related test questions were dropped from the final test since no between-group differences emerged in Experiment 1. Following the initial testing/filler activity phase, all participants listened to the post-event narrative. As in Experiment 1, there were 24 critical narrative statements: eight introduced consistent details, eight introduced neutral details, and eight introduced misleading details. During the narrative presentation, participants completed a same-modality distractor task adapted from Foerde, Knowlton, and Poldrack (2006). Importantly, after the narrator read each critical sentence, a high pitched tone signaled participants to listen to a series of eight high and low pitched tones. Participants counted the number of high pitched tones in the series, and recorded the number on a sheet. The number of high pitched tones following each signal randomly varied from one to eight. Importantly, prompts to count beeps were presented 1.5 seconds *after* the presentation of critical information, during the presentation of filler narrative information. This ensured that critical details were *encoded*, just could not be then preferentially processed. A pilot study confirmed this manipulation was sound. Eight undergraduates from Tufts University who did not complete any other experiment served as pilot participants. Participants first listened to the recording of the narrative. Presence of the beep counting task was manipulated within subjects. On half of the critical narrative trials, beeps were presented after critical details. On the other half no beeps were presented. After a brief 5-minute retention interval, participants completed a cued recall task. After the cued recall task, they completed a forced choice, two alternative recognition task. Participants did not differ in recall accuracy when beeps were present (M = .47) compared to when they were not present (M = .44), t = .571. Participants did not differ in recognition accuracy when beeps were present (M = .72) compared to when they were not present (M = .75), t = .407.

After the narrative phase, participants completed a final cued recall test identical to that taken in the initial test phase querying details from the video. They were allowed to withhold responses. In addition, confidence ratings for each response were collected for all responses as in previous experiments.

Experiment 3 Results

Memory Performance

Where appropriate, all pairwise comparisons used a Bonferroni correction. During the initial recall test, .62 of participants' responses were accurate and .06 produced misinformation spontaneously.

Accurate Video Recall on Final Test. Table 3 presents the accurate video recall probabilities on the final test. A 3 (Item Type: Consistent, Neutral, Misleading) x 2 (Test Group: Repeated, Standard) x 2 (Attention: Full, Divided) analysis of variance

(ANOVA) was conducted on final accurate recall. For the omnibus test, main effects are reported first, followed by interactions. A main effect of item type was significant, F (2, 240) = 48.91, $p < .001$, $\eta_p^2 = .29$. Participants were most accurate on consistent questions (M = .74) as compared to neutral (M = .58) and misleading question trials (M = .52). Follow up tests revealed consistent trials resulted in significantly greater accuracy as compared to neutral, t(123) = 7.59, $p < .01$, d = .68. In addition, participants were more accurate on neutral trials as compared to misleading trials, t(123) = 2.60, $p < .05$, d = .23. A main effect of attention was revealed, F (1, 120) = 4.20, $p < .05$, $\eta_p^2 = .03$. Participants in the divided attention group (M = .64) were more accurate across all trial types than participants in the full attention group (M = .58). There was not a main effect of testing group, $p = .13$.

The omnibus test revealed a significant item type by testing group interaction, F (2, 240) = 4.76, $p < .01$, $\eta_p^2 = .4$. The interaction was driven by the difference between testing groups on consistent trials. Repeated test participants were more accurate (M = .79) than standard participants (M = .69), t (122) = 2.83, $p < .05$, d = .51. Comparisons on neutral ($p = .08$) and misleading trials ($p = .43$) were not significant. A testing group by attention interaction was significant, F (1, 120) = 5.37, $p < .05$, $\eta_p^2 = .04$. In the full attention manipulation, there was no difference between testing groups in final test accuracy [Standard = .59; Repeated = .57, $p = .58$]. In the divided attention manipulation, the repeated test group was significantly more accurate than the standard group [Standard = .59; Repeated = .70, t (49.18) = 2.69, $p < .05$, d = .70].

Misinformation Production on Final Test. Average misinformation production is reported in Table 4. Further analyses assessed the effects of the testing and attention manipulations on the likelihood of responding with misleading details on the final memory test. These analyses were limited to misleading trials. In the Full Attention group, standard test participants reported misleading details on 24% of trials, and repeated test participants on 42% of trials, indicating a retrieval enhanced suggestibility effect. In the Divided Attention group, standard test participants reported misleading details on 20% of trials, and repeated test participants on 27% of trials. A 2 (Testing Group: Standard, Repeated) x 2 (Beep Counting: Present, Not Present) ANOVA on misinformation production revealed a main effect of testing group,

F (1, 120) = 9.07, $p < .01$, $\eta_p^2 = .07$. Participants in the repeated test group (M = .34) produced more misleading details than the standard group (M = .22). A main effect of attention was also observed, F (1, 120) = 5.74, $p < .05$, $\eta_p^2 = .05$. Participants in the full attention group (M = .33) produced more misleading details than participants in the divided attention group (M = .23). The interaction between testing group and attention was not significant, F (1, 120) = 1.89, $p = .17$.

Final Test Confidence. As in Experiment 1, separate analyses were conducted on confidence for final test questions answered with correct video details, and final test questions answered with incorrect misleading details. A 2 (Testing Group: Standard, Repeated) x 2 (Attention: Full, Divided) x 3 (Item Type: Consistent, Neutral, Misleading) mixed ANOVA with Testing Group and Attention as between subjects factors, and Item Type as a within subjects factor revealed a main effect of Item Type. Greenhouse-Geisser values are reported to correct for a violation of the assumption of sphericity, F (1.8, 207.5) = 4.28, $p < .05$, $\eta_p^2 = .04$. Follow up tests revealed that participants were more confident on consistent compared to neutral trials, t (121) = 3.47, *p* < .01, *d* = .31. The comparison between neutral and misleading (*p* = .24) and misleading and consistent (*p* = .11) were not significant. A three way interaction was also significant, *F* (1.8, 207.5) = 3.15, *p* = .05, η_p^2 = .03. After correcting for multiple comparisons none of the follow up tests were significant (all *p*s > .05). Mean confidence ratings are reported in Table 5.

Test 1 Confidence and Final Test Performance

These analyses were conducted in the same manner as in Experiment 1, however the addition of the between subjects factor of attention required separate 2-way ANOVA. Of a possible 1488 individual Test 1 confidence responses, 90% were valid.

Consistent Trials. A 2 (Attention: Full, Divided) x 2 (Test 2 Video Accuracy: Correct, Incorrect) ANOVA was conducted on mean confidence ratings on consistent trials. A main effect of accuracy was revealed, *F* (1, 161) = 4.07, *p* < .05, η_p^2 = .03. When correct responses were provided on Test 2, the associated Test 1 confidence ratings (*n* = 85; *M* = 32.1) were significantly lower than trials on which incorrect responses were provided on Test 2 (*n* = 80; *M* = 42.8). That is, in contrast to what would be predicted from a standard hypercorrection account, participants were more likely to correct low-confidence errors. The main effect of attention (*p* = .92) and the attention by test 2 accuracy interaction (*p* = .15) were not significant.

Misleading Trials. More relevant to RES are the analyses on misleading trials. First, a 2 (Attention: Full, Divided) x 2 (Test 2 Misinformation Production: Produced, Not Produced) was conducted on mean confidence ratings on misleading trials where Test 1 was answered incorrectly. A main effect of production was revealed, *F* (1, 147) = 3.99, *p* < .05, η_p^2 = .03. When misleading details were produced on Test 2, the associated

Test 1 confidence ratings (*n* = 89; *M* = 30.9) were significantly lower than trials on which misinformation was not produced on Test 2 (*n* = 60; *M* = 43.1). Thus, participants were more likely to endorse misinformation when they had reported a test 1 error with low confidence. The main effect of attention (*p* = .10) and the attention by Test 2 accuracy interaction (*p* = .76) were not significant.

On misleading trials where test 1 questions were answered correctly, a 2 (Attention: Full, Divided) x 2 (Test 2 Misinformation Production: Produced, Not Produced) was conducted on mean Test 1 confidence ratings. A main effect of production was revealed, *F* (1, 283) = 6.48, *p* = .01, η_p^2 = .02. When misleading details were produced on Test 2, the associated Test 1 confidence ratings (*n* = 59; *M* = 65.0) were significantly lower than trials on which misinformation was not produced on Test 2 (*n* = 228; *M* = 88.5). Thus, participants were more likely to endorse misinformation when they had correctly answered Test 1 questions with low confidence. The main effect of attention (*p* = .59) and the attention by Test 2 accuracy interaction (*p* = .65) were not significant.

Discrepancy Recollection

A 2 (Testing Group: Standard, Repeated) x 2 (Attention: Full, Divided) ANOVA assessed participants' likelihood of detecting a discrepancy between details presented in the video and narrative on misleading trials. A marginal effect of group was found, *F* (1, 114) = 3.60, *p* = .06. Participants in the repeated test group accurately remembered discrepancies on 61% of trials, while the standard group remembered discrepancies on 54% of trials. The effect of attention (*p* = .72) and the testing group by attention interaction (*p* = .19) were not significant. Means are reported in Table 6.

Experiment 3 Discussion

Overall, participants in Experiment 3 were most accurate on final test questions associated with consistent trials, and repeated test participants even more so than standard test participants. Consistent with previous research, RES was demonstrated with final test misinformation production but not final test accuracy (Thomas et al., 2010; Gordon & Thomas 2014; Thomas et al., 2015). A difference was observed on final test accuracy, however, between *attention* groups. Full attention participants were less accurate than divided attention participants on the final test. On the surface, this may seem counterintuitive. Full attention should support memory. However, the attention manipulation occurred during narrative learning, and the final test assessed original event memory from the video. Thus, when full attention was directed to narrative processing, original event memory was impaired as compared to when attention to narrative processing was divided. If dividing attention decreases the likelihood of responding on the final test with inaccurate misleading details, and this accompanies an increase in reports of accurate details, then overall accuracy should be counter intuitively higher under divided as compared to full attention.

Turning to the production data, a standard RES effect was demonstrated. The repeated test group reported more misleading details than the standard group. In addition, participants in the full attention group reported more misleading details compared to the divided attention group. If the beep counting task effectively disrupted individuals' ability to differentially attend to, or spend more time thinking about, misleading narrative details after listening to them in the narrative, then this finding suggests that attention may be important to eyewitness suggestibility. The strongest support for this argument would have been found in a testing group by attention interaction demonstrating RES in the full attention group, but not the divided attention group. However, this interaction was not significant.

Finally, the Test 1 confidence analyses replicated the findings from Experiment 1. A standard account of hypercorrection was not supported in the analysis on consistent trials. More relevant to RES, participants were more likely to endorse misinformation when they were correct on Test 1, and those correct Test 1 responses were associated with low confidence, then when such responses were associated with higher confidence.

Experiment 4

Experiment 4 was designed to examine the role of attention in forward testing effects. Further, it sought to provide additional support for the categorization of RES as an instance of test-potentiated learning. As in Experiment 3, the primary manipulation in Experiment 4 was disruption of participants' ability to preferentially process, or attend to, critical narrative details. Learning of narrative details was directly measured on a final memory test.

Experiment 4 Method

Participants

A total of 128 undergraduates from Tufts University participated and were compensated either with course credit or $15.

Materials and Procedure

The materials and procedure were identical to Experiment 3 with one change.

One the final test, participants were instructed to report what they remembered learning in the narrative only. They were allowed to omit responding.

Experiment 4 Results

Memory Performance

During the initial recall test, .63 of participants' responses were accurate and .07 produced misinformation spontaneously.

Final Test Narrative Recall. Average narrative recall accuracy is reported in Table 7. On the final test, participants were instructed to report information they learned in the narrative only. For consistent trials, this was a video detail and thus *repeated* information in the narrative. For misleading trials, this was a detail presented only in the narrative and thus *new* information in the narrative. Neutral trials did not provide specific details in the narrative so were excluded from analysis. Recall of repeated (consistent) and new (misleading) details were examined in separate analyses. Follow up tests for each analysis used a Bonferroni correction unless otherwise stated.

Consistent Trials. A 2 (Testing Group: Standard, Repeated) x 2 (Attention: Full, Divided) ANOVA examined recall of details that were presented in the video and repeated in the narrative. A main effect of testing group was revealed, F (1, 124) = 8.61, $p < .01$, $\eta_p^2 = .07$. Participants in the repeated test group (M = .88) recalled more consistent details from the narrative than participant in the standard group (M = .80). A main effect of attention was also revealed, F (1, 124) = 7.18, $p < .01$, $\eta_p^2 = .06$. Participants in the full attention group (M = .88) recalled more consistent details from the narrative than participants in the divided attention group (M = .80). Finally, the group by attention interaction was significant, F (1, 124) = 9.57, $p < .01$, $\eta_p^2 = .07$. In the full attention group, there was no difference in narrative recall for consistent items between the standard (*M* = .88) and repeated (*M* = .87) groups, *t* = .129. In the divided attention group, the repeated test group (*M* = .88) recalled more details than the standard group (*M* = .72), *t* (46.71) = 3.77, *p* < .01, *d* = .99.

Misleading Trials. A 2 (Testing Group: Standard, Repeated) x 2 (Attention: Full, Divided) ANOVA examined recall of misleading details that were only presented in the context of the narrative. A main effect of testing group was revealed, *F* (1, 124) = 66.60, $p < .001$, $\eta_p^2 = .35$. Participants in the repeated test group (*M* = .78) recalled more misleading details from the narrative than participant in the standard group (*M* = .47). A main effect of attention was also revealed, *F* (1, 124) = 20.35, $p < .001$, $\eta_p^2 = .14$. Participants in the full attention manipulation (*M* = .71) recalled more misleading details from the narrative than participants in the divided attention manipulation (*M* = .55). The interaction effect was marginal, *p* = .06.

Final Test Confidence. A 2 (Testing Group: Standard, Repeated) x 2 (Attention: Full, Divided) ANOVA on mean final test confidence ratings for accurately reported consistent narrative details revealed a main effect of Testing Group, *F* (1, 124) = 20.64, $p < .001$, $\eta_p^2 = .14$. The repeated test group was more confident that the standard test group on these trials. The effect of attention was not significant (*p* = .25) nor was the testing group by attention interaction (*p* = .79).

A 2 (Testing Group: Standard, Repeated) x 2 (Attention: Full, Divided) ANOVA on mean final test confidence ratings for accurately reported misleading narrative details also revealed a main effect of Testing Group, *F* (1, 119) = 18.69, $p < .001$, $\eta_p^2 = .14$. The repeated test group was more confident that the standard test group on these trials. The effect of attention was not significant (*p* = .66) nor was the testing group by attention interaction (*p* = .14).Mean confidence ratings are reported in Table 8.

Discrepancy Recollection

A 2 (Testing Group: Standard, Repeated) x 2 (Attention: Full, Divided) ANOVA assessed participants' likelihood of recollecting a discrepancy between details presented in the video and narrative on misleading trials. One participant was dropped from this particular analysis for failure to comply with the discrepancy recollection instructions. A main effect of group was found, F (1, 123) = 6.86, $p < .01$, $\eta_p^2 = .05$. Participants in the repeated test group (*M* = .62) were more likely to accurately recollect a discrepancy than participants in the standard group (*M* = .52). A main effect of attention was also revealed, F (1, 123) = 10.68, $p < .001$, $\eta_p^2 = .08$. Participants in the full attention group (*M* = .63) were more likely to accurately recollect a discrepancy than participants in the divided attention manipulation (*M* = .51). The interaction was not significant, *p* = .2. Means are reported in Table 6.

Experiment 4 Discussion

Results from Experiment 4 demonstrated that the repeated test group recalled more consistent and misleading narrative details compared to the standard group when the final test directly queried narrative details. In addition, repeated testing and full attention at narrative encoding independently increased participants' ability to recollect discrepancies between the video and narrative on misleading trials. That is, testing promoted learning of narrative details, even when the ability to review or rehearse those details was interrupted. Likewise, attention promoted learning even in the absence of testing.

Finally, the results indicate that test-enhanced attention is not necessary for learning of new information. This is qualified by the absence of an attention by testing interaction on misleading narrative detail recall. That is, even in situations when attention was divided, repeated testing continued to result in potentiated learning.

General Discussion

The aim of this dissertation was two-fold: 1) to carefully evaluate the proposal that retrieval enhanced suggestibility is an example of test-potentiated learning, and 2) to examine whether immediate testing increases learning and reporting of new information, including misinformation, via an attention mechanism. The results across experiments indicate that attention to post-event information may be important to enhanced suggestibility, and that testing may be one way to promote attention. In contrast, attention is not a necessary component to test-potentiated learning of new information. These primary findings are elaborated upon and implications for broader theories of test enhanced learning are addressed in the general discussion.

Attention and Enhanced Suggestibility

Previous research has suggested that immediate testing may affect attention allocation to post-event information in the misinformation paradigm. Changes in attention allocation affect the encoding and accessibility of misleading post-event details, and result in enhanced suggestibility (Gordon & Thomas, 2014; Gordon, Thomas, et al., 2015; Thomas et al., 2010). Experiments in this dissertation extended this previous work and provided a strong test of the attention allocation hypothesis by directly examining the contribution of attention to enhanced processing and learning of post-event information.

Collectively, the results across experiments show some support for the role of attention in enhanced suggestibility. In Experiment 1, there was no between-testing group difference in reaction times to a secondary probe task designed to measure attention. However, evidence from the attention literature suggests that the secondary probe task may have *increased* attention in the non-tested group (Jiang & Swallow, 2010; Mulligan et al., 2014; Spataro et al., 2013). While there was no direct evidence in Experiment 1 that attention was augmented in either group, prior studies suggest that repeating testing *does* differentially impact attention to post-event details (Gordon, Cernasov, et al., 2015; Gordon & Thomas, 2014; Gordon, Thomas et al., 2015). Further, in the present Experiment 1, *both* groups demonstrated enhanced misinformation susceptibility. As highlighted in the discussion on the first experiment, the standard test group reporting misleading details on 41% of trials, compared to the 27% typically observed in similar studies (e. g., Chan et al., 2009; Thomas et al., 2010). Thus, the standard group demonstrated enhanced suggestibility, in the absence of initial retrieval.

Experiment 3 attempted to disrupt attention, as operationalized as differential processing or reviewing of narrative details are they are encoded. Suggestibility, as measured by production, was greater overall when attention was not disrupted. RES, as measured by production, was also observed, but did not change as a function of attention disruption. This suggests, in contrast to Experiment 1, that attention may be less important to suggestibility. However, it is important to note that the observed power on the test of the interaction between testing group and attention was relatively low (.28). Further, the numeric difference between average production in the repeated test/full attention group (M = .42) and the repeated test/divided attention group (M = .27) was in the predicted direction. While this indicates that the test was likely underpowered, the conclusion at this point can only be that attention cannot be the primary factor underlying retrieval enhanced suggestibility.

Despite the inconclusive findings from Experiments 1 and 3, it is still important to question whether retrieval is necessary for ‘retrieval enhanced suggestibility’. Attention may be the more important factor, and testing may be just one way to effectively promote attention. Unpublished data from Gordon, Cernasov, et al. (2015) support this conclusion. In this study, participants viewed a video of a fictional crime and were then randomly assigned to one of three groups. A repeated test group took an immediate test on the event. A standard group completed a filler activity. A third group, the exogenous attention group, also completed a filler activity. Next, each group read a narrative that included consistent, neutral, and misleading details about the event. Importantly, the repeated test and standard groups were presented the narrative in normal type face. The attention group, however, was presented a modified narrative in which critical event details were presented in red, underlined font. On a final memory test presented immediately after the narrative phase, both the repeated test *and* attention groups demonstrated equivalent levels of enhanced suggestibility compared to the standard group. Interesting, when a 48-hour delay was imposed between the narrative and the final test, enhanced suggestibility disappeared and the repeated test group was *more* accurate on misleading trials compared to the standard and attention groups who did not differ. Attention and testing both influenced suggestibility, but when fluency, or the ease with which information comes to mind at test, was minimized with a delay, better memory for the original event, even in the context of post-event interference, was demonstrated only for participants who took an immediate test before new learning. These results are consistent with the Experiments 2 and 4 of the present study, and provide evidence for the disassociation between attention and testing.

Gordon and colleagues argued that misleading post-event narrative details are learned better after testing. Test-enhanced learning occurs because tested participants shift processing resources to misleading details in the narrative (Gordon & Thomas, 2014; Gordon et al., 2015). Increases in attention lead participants to respond with misinformation on later memory tests, because the ease with which test-potentiated narrative learning comes to mind, or retrieval fluency, biases responding even if original event details are still available in memory (cf., Baddeley, 1982; Jacoby & Dallas, 1981). The present experiments provide some support for the view that RES results from a response bias at final test. While testing certainly enhances a deeper learning of misinformation (Experiments 2 and 4) relative to non-tested group, response biases that manifest as enhanced suggestibility in the present experiments have some relation to attention. It is possible that differentially attending to misleading narrative details temporarily increases accessibility to these details in memory. Individuals are biased to report these fluently retrieved, yet inaccurate details, on a final memory test. Retrieval fluency was not measured directly in the present experiment. However, a pattern of data show support for this interpretation. Although the critical analysis was non-significant, when fluency was tempered by the minimization of attention in Experiment 3, there was a numerical reduction in enhanced suggestibility in the repeated test group. This pattern is consistent with studies that have shown when participants are encouraged to override fluent responding, RES effects disappear and access to originally learned details increases (Gordon, Cernasov, et al., 2015; Gordon & Thomas, 2014; Thomas et al., 2010).

The probe task used to measure attention in Experiments 1 and 2, and the beep counting task used to disrupt differential processing in Experiments 3 and 4 were novel tasks in the context of retrieval enhanced suggestibility studies. The data suggest that while the probe task may have measured attention, it also captured attention, and in Experiment 1similarly enhanced suggestibility compared to a repeated test manipulation. While this is an interesting hypothesis and supported by findings from a broader literature on attention (Mulligan, Spataro, & Picklesimer, 2014; Spataro, Mulligan, & Rossi-Arnaud, 2013; Swallow & Jiang, 2010), it is admittedly post-hoc, and this possibility should be explored in future studies. In addition, it is possible that the beep counting task (Experiments 3 and 4) did not fully disrupt participants' ability to review critical narrative details. While pilot testing indicated that the task did not reduce participants' ability to learn and report narrative details, it is difficult to be sure where encoding ends and 'extra' processing begins. Further, while a within-subjects design, pilot testing was limited to only eight participants. Additional work will be needed to further disentangle these processes in context of RES.

Revisiting Alternative RES Theories

Two accounts have been proposed to explain RES: Retrieval Fluency and Reconsolidation. The strongest evidence for the reconsolidation account of RES was demonstrated in Chan and LaPaglia (2013) who employed a source-free final recognition memory test in a standard RES paradigm. This procedure presented participants with details from either the original event or the narrative, and required them to respond "old" if they remembered the information from *either* source, and to respond “new” otherwise. They reasoned that this procedure would rule out the possibility that RES results of bias from factors such as source confusion. If repeated test participants showed reduced accuracy on final tests after exposure to misleading details because they simply were unsure whether a detail was learned during the original learning event (video) or post-event learning (narrative), they should respond “old” on the source-free test when presented with details from the video. Chan and LaPaglia found instead that repeated test participants responded that these items were “new”. They argued that memory accessibility could be measured regardless of the presence of source, and concluded that original memories are inaccessible when immediate testing is followed by misinformation.

The present findings contrast with those of Chan and LaPaglia, using their same conceptualization of accessibility. That is, the discrepancy recollection procedure did not require explicit source memory, but simply required acknowledgement of two separate memory traces. In Experiments 2 and 4, repeated test participants were better able to recollect a discrepancy between video and narrative content on misleading trials than participants who were not tested. If testing followed by misinformation exposure eliminates an original memory discrepancy recollection would not be possible. Methodological differences exist between the present study and Chan and LaPaglia. In the present Experiments 2 and 4, repeated test participants took an immediate test on the video *and* a final cued-recall test on narrative details prior to the discrepancy recollection task. Repeated test participants in the Chan and LaPaglia study only took an immediate test on the video prior to the source-free recognition task. It is presently unknown whether the addition of a more demanding retrieval task prior to the source free recognition task would change the pattern of results observed in Chan and LaPaglia, but is a question for future research.

Another explanation for RES was explored in this dissertation for the first time. It was possible that participants, after testing, used feedback from the narrative to ‘correct’ responses on Test 1, and that this corrective behavior was contingent upon accuracy in Test 1 responses. This explanation of RES was not supported. On the contrary, participants were more likely to report misleading details when Test 1 answers were associated with low confidence. This is important to understanding retrieval enhanced suggestibility effects. As confidence in memory can change as a function of the number of times information is retrieved from memory (c. f., Thomas et al., 2015), it is important to understand the types of memory representations that are most susceptible to interference. The lack of evidence for a hypercorrection account of RES in consistent with other studies that have explored RES as a more general feedback effect. While Test 1 confidence was not examined, Gordon and Thomas (2014) proposed that perhaps after taking an initial test, participants simply used narrative details feedback for their performance. Regardless of what is reported on Test 1, in the case of misleading trials narrative details will always contradict Test 1 responses. They argued that perhaps participants simply used the misleading narrative details as corrective feedback. To test this hypothesis, they included a manipulation in which Test 1 questions could not be answered by the narrative. Test 1 questions instead were episodically related to final test questions, which *could* be answered by narrative details. Essentially, in this manipulation

Test 1 feedback was not possible. They found that even in the no-feedback group, RES was observed.

Revisiting Theories of Forward Effects of Testing

The results of Experiments 2 and 4 revealed that after immediate testing, individuals learn subsequently presented information better. Repeated test participants were more likely to recall new, misleading narrative details on a final test of memory compared to participants who did not take an immediate test. These experiments were important for two reasons. First, they provided a direct demonstration of test-potentiated learning in the RES paradigm. Second, directly measuring post-event learning afforded a novel perspective in examining theories of test-enhanced learning. Theories of test potentiated learning, or forward effect of testing, can be divided into two basic categories: retrieval and encoding explanations. I predicted an encoding explanation of test-potentiated learning. Encoding explanations propose that testing changes participants' encoding strategy, enhancing attention to and encoding of post-test information informative (e. g., Reynolds & Anderson, 1982; Wissman et al., 2011).

Prior studies indicated that testing influenced attention during new learning, as measured by reading times of sentences in the post-event narrative, in the RES paradigm (Gordon & Thomas, 2014; Gordon, Thomas, et al., 2015). If RES is best categorized as an instance of test-potentiated learning, as suggested by these same studies, I hypothesized that attention should also be important to test-potentiated learning. Previous research examining attention and test-potentiated learning manipulated attention by changing features of studied material such as font color and size (e. g., Shapiro & Gordon, 2012; Richland et al., 2009). Using these kinds of manipulations, attention was always found to be inferior to retrieval practice. That is, taking a test led to greater learning than comparative attention manipulations. The findings of Experiments 2 and 4 are consistent with these studies, and contrary to my original hypothesis. In Experiment 2, when the probe task increased attention to the post-event narrative even in the standard test group, learning of the narrative was still superior in the repeated test group. Moreover, Experiment 4 demonstrated test-potentiated learning of new narrative details even when any extra attention to these details was disrupted by a beep-counting task. Thus, consistent with previous studies discounting test-potentiated learning as an attention effect, an encoding account of test-potentiated learning in the misinformation paradigm was not supported in the present study.

In contrast to an encoding account, retrieval explanations of testing effects assume that testing encourages differentiation between learning episodes. When initially learned information is tested, it is strengthened (e. g., Roediger & Karpicke, 2006). This process counteracts proactive interference, or interference from originally learned information, by segregating source information from the original learning episode and subsequent learning episodes (e.g., Bäuml and Kliegl, 2013; Pastötter et al., 2011; Szpunar et al., 2008; Weinstein, Gilmore, Szpunar, & McDermott, 2014). On later tests, more information can be recalled from subsequent learning episodes because individuals can use source information as a retrieval cue, and are better equipped to accurately monitor sources. The resulting memory performance manifests as test-potentiated learning or forward effects. A majority of the evidence supporting this view has been gathered in verbal learning studies. The present study extended this to more complex materials. Importantly, the pattern of results across the present experiments supports a retrieval-based explanation of test-potentiated learning of new, misleading information. First, in both Experiments 2 and 4, participants who took an initial test recalled more new information from the narrative than participants who were not tested. Further, both experiments revealed that the repeated test participants recalled more discrepancies between video and narrative content than the standard group. Accurate discrepancy recollection suggests source segregation.

Wahlheim (2015) suggests that both source segregation *and* source integration can reduce proactive interference and promote learning. In his experiment, participants studied a list of A-B word pairs (L1). They then either restudied the word pairs, or were tested on them. Next, participants studied a second list of pairs (L2). The second list consisted of A-B pairs, A-D pairs (same cue, new target) and C-D pairs (completely new). On a final cued-recall test, participants recalled L2 responses and then indicated when they recollected that responses had earlier changed between lists. Most relevant to the present study, and forward effects of testing, is participants' recall of A-D pairs. Wahlheim found that 1) testing promoted recollection of change, and 2) participants recalled more A-D pairs when change was recollected than when change was not recollected. These results mimic those found in the present study.

Revisiting Theories of Backward Effects of Testing

In addition to investigating forward effects of testing, a simple exploration of the elaborative retrieval hypothesis was included in Experiment 1 to look at one potential way testing may promote backward effects. ERH (Carpenter, 2009, 2011; Carpenter & DeLosh, 2006) proposes that presentation of a retrieval cue activates information in memory that is related to the cue during the search for the target. The additional activated information is encoded along with the cue-target information, providing additional retrieval routes through which to access the target on later memory tests. In the present Experiment 1, questions that targeted video details that were *related* to those on initial test questions were included on the final memory test. It was hypothesized that if backward effects of testing occur due to an elaboration mechanism, then after initial testing memory performance for the ‘related’ final test questions would increase. Repeated and standard test participants performed equally on final test related questions, which suggests that elaboration is not important to testing effects. This finding, however, conflicts with other studies that demonstrated testing promotes recall of non-tested material (Chan et al., 2006; Rowland & DeLosh, 2014). It is important to note that in Experiment 1, between-testing group differences were also not observed on misinformation production or discrepancy recollection, perhaps because the secondary probe task presented in this experiment boosted attention to narrative content even in the standard test group. It is unclear the extent to which the probe manipulation may have impacted elaborative retrieval. A limitation of the present study was dropping related test questions from Experiment 3, as this explanation could have been more carefully evaluated there. Thus, the present results cannot rule out the elaborative retrieval hypothesis as a viable explanation of backward effects of testing.

Future Directions

The results of the present experiments not only informed the primary goals of the study, but led to new questions that can be addressed in future work. First, attention and fluency bias explanations of enhanced suggestibility have been supported independently (e. g., Gordon & Thomas, 2014; Gordon, Thomas, et al., 2015; Thomas et al., 2010). The present study presents some evidence for attention as a factor in enhanced suggestibility. The perspective taken is that attention may impact temporary accessibility of misleading details leading to a retrieval fluency bias during a final memory test. Direct support for the link between attention and retrieval fluency, however, is currently lacking. It is important to better-specify this relationship in future work.

Second, the pattern of results across experiments point to the relationship between discrepancy recollection, source discrimination, and retrieval fluency. In the present study accurate discrepancy recollection may measure source discrimination. Effective source discrimination may indicate that fluency-based responding was overridden. This is consistent with previous work showing that fluency is less-relied upon when source discrimination is encouraged (e. g., Thomas et al., 2010). While in the present study discrepancy recollection was measured at the end of the experimental procedure, it is possible that discrepancies are detected by participants earlier. This could impact how information is processed, and the ease with which it is accessed on subsequent memory tests. Consistent with this proposal, Gordon, Thomas, et al. (2015) found that repeated test participants spent the most time reading misleading narrative details when they directly contradicted with what they reported on Test 1. Increased reading times predicted later misinformation production on the final test. A direct examination of the relationship between discrepancy recollection and fluency can further the encoding account of enhanced suggestibility.

Conclusions

This dissertation was built upon the premise that RES is an indirect example of the forward effect of testing. Testing certainly promotes learning of new details in the present paradigm, and these new details are misleading details. Suggestibility, or the tendency to report these inaccurate details when they are not directly queried, is primarily dependent on the accumulation of retrieval fluency. One way retrieval fluency of misleading details may accumulate is when they are differentially attended to. Testing is one way to increase attention. When a manipulation designed to minimize the contributions of increased attention was present, fluent retrieval of misleading details numerically decreased. Yet, when narrative learning was directly queried enhanced learning was still observed after testing. This suggests that while attention may be a catalyst for enhanced suggestibility, it may not be necessary for test-potentiated learning. Thus, results from this dissertation provide some support for the attention allocation hypothesis of RES. However, RES can no longer be considered as primarily resulting from test-potentiated learning of post-event misinformation. Rather, as discussed earlier, retrieval fluency bias, perhaps provoked by attention shifts, may be primarily responsible for the effect. When the retrieval fluency bias is overridden through the careful evaluation of source at final test (e. g., Gordon & Thomas, 2014; Thomas et al., 2010) enhanced suggestibility disappears and memory accuracy is augmented. The present findings have important implications for eyewitness memory. When their attention is drawn to inaccurate details about an event, perhaps after answering questions at the scene of the crime, eyewitnesses may be more likely to report these inaccuracies when they later recount the crime. However, when appropriate retrieval cues are presented that encourage careful source discrimination, eyewitnesses can successfully retrieve the correct event details.

References

Arnold, K.M., & McDermott, K.B. (2013). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review, 20*, 507-513.

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review, 89,* 708–729.

Balota, D. A., Black, S. R., & Cheney, M. (1992). Automatic and attentional priming in young and older adults: A reevaluation of the two process model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 485-502.

Bäuml, K.-H. T. & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language, 68*, 39-53.

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23,* 1059–1082.

Bekerian, D. A., & Bowers, J. M. (1983). Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 139–145.

Belli, R. F. (1993). Failure of interpolated tests in inducing memory impairment with final modified tests: Evidence unfavorable to the blocking hypothesis. *The American Journal of Psychology*, *106*, 407–427.

Bowman, L. L., & Zaragoza, M. S. (1989). Similarity of encoding context does not influence resistance to memory impairment following misinformation. *The American Journal of Psychology*, *102,* 249–264.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1491–1494.

Butterfield, B., & Metcalfe, J. (2006). The correction of errors held with high confidence. *Metacognition Learning, 1,* 69-84.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35,* 1563-1569.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37,* 1547-1552.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268-276.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642.

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General, 116,* 38–49.

Chan, J. C. K., & Langley, M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 248-255.

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*, 418-432.

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

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 533-571.

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*, 66-73.

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

Christiaansen, R. E., & Ochalek, K. (1983). Editing misleading information from memory: Evidence for the coexistence of original and postevent information. *Memory & Cognition*, *11*, 467–475.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology, 17*, 177-184.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences, USA, 103*, 11778–11783.

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science, 20,* 20-23.

Gates, A.I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *6*(40).

Gordon, L. T., Cernasov, P., Bulevich, J. B., & Thomas, A. K. (2015). Endogenous and exogenous cuing in episodic memory. in preparation.

Gordon, L. T., & Shapiro, A. M. (2012). Priming correct information reduces the misinformation effect. *Memory & Cognition*, *40,* 717-726.

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

Gordon, L. T., Thomas, A. K., & Bulevich, J. B. (2015). Looking for Answers in All the Wrong Places: How Testing Facilitates Learning of Misinformation. *Journal of Memory and Language*, in press.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40*, 505-513.

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 290-296.

Henkel, L.A., Johnson, M.K., & De Leonardis, D.M. (1998). Aging and source monitoring: cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127,* 251-268.

Innocence Project. *Facts on Post-Conviction DNA Exonerations.* Retrieved November 10, 2013 from http://www.innocenceproject.org/ Content/Facts_on_ PostConviction_DNA_ Exonerations.php#

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, 18,* 879–919.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 112-127.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Kang, S. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology, 19,* 528-558.

Karpicke, J. D. & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language, 67*, 17-29.

Kelley, C. M., & Lindsay, S. D. (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.

Kinchla, R. A., (1992). Attention. *Annual Review of Psychology, 43,* 711-742.

Knowles, W. B. (1963). Operator loading tasks. *Human Factors, 5,* 151-161.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 989-998.

Kroll, N. E., Ogawa, K. H., & Nieters, J. E. (1988). Eyewitness memory and the importance of sequential information. *Bulletin of the Psychonomic Society, 26,* 395–398.

Larsen, D. P., Butler, A. C., & Roediger, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: a randomized controlled trial. *Medical Education, 43*, 1174-1181.

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.

Loftus, E. F., Schooler, J. W., & Wagenaar, W. (1985). The fate of memory: Comment on McCloskey and Zaragoza. *Journal of Experimental Psychology: General*, 114, 375–380.

Lynch, J. G., & Srull, T. K. (1982). Memory and attentional factors in consumer choice: Concepts and research methods. *Journal of Consumer Research, 9,* 18-37.

McCloskey, M. & Cohen, N. (1989) Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (ed.) *The Psychology of Learning and Motivation,24,* 109-164

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

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 168–180.

McDaniel, M., Anderson, J., Derbish, M., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*, 494–513.

Moray, N. (1967). Where attention is limited: A survey and model. *Acta Psychologica, 27*, 84-92.

Müller, G. E., & Pilzecker, A. (1900).Experimentelle Beiträge zur Lehre vom Gedächtniss. Zeitschrift für Psychologie. *Ergänzungsband*, *1,* 1–300.

Mulligan, N. W., Spataro, P., & Picklesimer, M. (2014). The attentional boost effect with verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1049-1063.

Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews Neuroscience, 10*, 224-234.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.). *The psychology and learning and motivation* (Vol. 26, pp. 125–173). New York: Academic Press.

Pashler, H. (1998). *The Psychology of Attention.* MIT Press: Cambridge, MA.

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37*, 287-297.

Postman, L. (1962). Retention of first-list associations as a function of the conditions of transfer. *Journal of Experimental Psychology*, *64,* 380-387.

Postman, L., & Keppel, G. (1977). Conditions of cumulative proactive inhibition. *Journal of Experimental Psychology: General, 106,* 376-403.

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

Reynolds, R. E., & Anderson, R. C. (1982). Influence of questions on the allocation of attention during reading. *Journal of Educational Psychology*, *74,* 623-632.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15,* 243-257.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20-27.

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

Spataro, P., Mulligan, N. W., & Rossi-Arnaud, C. (2011). Attention and implicit memory. The role of the activation of multiple representations. *Experimental Psychology, 58*, 110-116

Stolz, J. A. (1996). Exogenous orienting does not reflect an encapsulated set of processes. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 187–201.

Swallow, K.M., & Jiang, Y.V. (2010). Attentional Boost Effect: Transient Increases in Attention to One Task Enhance Performance in a Second Task. *Cognition, 115,* 115-132.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392–1399.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49,* 83–90.

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.

Thomas, A. K., Chen, C., Gordon, L. T., & Tenbrink, T. (2015). Choose your words wisely: What verbal hesitation indicates about eyewitness accuracy. in revision.

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 Behavior, 13,* 181–193.

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition*, *43,* 27-38.

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1039-1048.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11,* 571-580.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50,* 449-455.

Wilford, M. M., Chan, J. C. K., & Tuhn, S. J. (2013). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. *Journal of Experimental Psychology: Applied.*

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011).The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18,* 1140-1147.

Wixted, J. T. (2004). The Psychology and neuroscience of forgetting. *Annual Review of Psychology, 55,* 235–269.

Yantis, S., & Jonides, J. (1984). Abrupt onsets and selective attention: Evidence from visual search. Journal of Experimental Psychology: *Human Perception and Performance, 10,* 601–620.

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.

Table 1

Log-Transformed Average Reaction Times to Probe Task (Standard Error in Parentheses)as a Function of Trial Type and Testing Group

	Consistent	Neutral	Misleading
Exp 1			
Standard	6.54 (.26)	6.58 (.30)	6.59 (.26)
Repeated	6.53 (.21)	6.55 (.21)	6.60 (.26)
Exp 2			
Standard	6.48 (.14)	6.51 (.17)	6.50 (.19)
Repeated	6.48 (.26)	6.44 (.24)	6.50 (.28)

Table 2

Median Reaction Times to Probe Task in Milliseconds (Standard Error in Parentheses)as a Function of Trial Type and Testing Group

	Consistent	Neutral	Misleading
Exp 1			
Standard	672 (203)	725 (279)	695 (181)
Repeated	650 (146)	677 (163)	676 (148)
Exp 2			
Standard	632 (96)	656 (104)	653 (118)
Repeated	639 (188)	625 (177)	657 (190)

Table 3

Average Proportion of Video Details Recalled Correctly on the Final Test in Experiments 1 and 3 as a Function of Item Type and Group. (Standard Error in Parentheses).

	Consistent	Neutral	Misleading
Experiment 1			
Standard	.79 (.04)	.52 (.03)	.45 (.04)
Repeated	.79 (.03)	.56 (.03)	.39 (.05)
Experiment 3			
Standard/Full	.72 (.04)	.53 (.04)	.52 (.04)
Repeated/Full	.75 (.04)	.56 (.04)	.40 (.04)
Standard/Divided	.66 (.04)	.55 (.04)	.55 (.04)
Repeated/Divided	.83 (.03)	.66 (.03)	.60 (.04)

Table 4

Average Misinformation Production on Misleading Trials on the Final Test in Experiments 1 and 3 as a Function of Group. (Standard Error in Parentheses).

Experiment 1	
Standard	.41 (.05)
Repeated	.47 (.05)
Experiment 3	
Standard/Full	.24 (.04)
Repeated/Full	.42 (.04)
Standard/Divided	.20 (.04)
Repeated/Divided	.27 (.04)

Table 5

Final Test Average Confidence Ratings (SD in Parentheses) When Final Test Queried the Video. Means Presented as a Function of Group, Trial Type, and Final Test Response Type. Scores Ranged from 0 (Not at all Confident) to 100 (Completely Confident).

Accurate Video Response on Final Test			
	Consistent	Neutral	Misleading
Exp 1			
Standard	82 (14)	77 (18)	84 (14)
Repeated	86 (16)	80 (20)	82 (27)
Exp 3			
Standard/Full	86 (11)	80 (20)	86 (15)
Repeated/Full	84 (16)	84 (15)	79 (20)
Standard/Divided	84 (12)	77 (18)	78 (18)
Repeated/Divided	86 (9)	82 (12)	87 (10)
Misleading Detail Response on Final Test			
	Consistent	Neutral	Misleading
Exp 1			
Standard	--	--	79 (21)
Repeated	--	--	74 (27)
Exp 3			
Standard/Full	--	--	70 (23)
Repeated/Full	--	--	65 (29)
Standard/Divided	--	--	63 (27)
Repeated/Divided	--	--	73 (20)

Table 6

Average Proportion of Discrepancies between Video and Narrative Reported as a Function of Item Type and Group. (Standard Error in Parentheses).

	Consistent	Neutral	Misleading
Experiment 1			
Standard	.23 (.03)	.36 (.04)	.53 (.05)
Repeated	.25 (.02)	.39 (.05)	.50 (.04)
Experiment 2			
Standard	.24 (.03)	.42 (.03)	.47 (.04)
Repeated	.22 (.03)	.32 (.05)	.66 (.04)
Experiment 3			
Standard/Full	.24 (.03)	.28 (.04)	.57 (.04)
Repeated/Full	.24 (.03)	.27 (.05)	.60 (.04)
Standard/Divided	.29 (.03)	.37 (.03)	.51 (.04)
Repeated/Divided	.20 (.03)	.22 (.04)	.64 (.04)
Experiment 4			
Standard/Full	.22 (.03)	.36 (.04)	.60 (.03)
Repeated/Full	.18 (.03)	.43 (.06)	.65 (.04)
Standard/Divided	.23 (.03)	.41 (.04)	.44 (.04)
Repeated/Divided	.18 (.03)	.39 (.05)	.58 (.04)

Table 7

Average Proportion of Narrative Details Recalled Correctly on the Final Test in Experiments 2 and 4 as a Function of Item Type and Group. (Standard Error in Parentheses).

	Consistent	Misleading
Experiment 2		
Standard	.71 (.04)	.43 (.05)
Repeated	.87 (.03)	.71 (.05)
Experiment 4		
Standard/Full	.88 (.02)	.59 (.04)
Repeated/Full	.87 (.03)	.83 (.03)
Standard/Divided	.72 (.04)	.35 (.04)
Repeated/Divided	.88 (.02)	.73 (.03)

Table 8

Final Test Average Confidence Ratings (SD in Parentheses) When Final Test Queried the Narrative. Means Presented as a Function of Group, Trial Type, and Final Test Response Type. Scores Ranged from 0 (Not at all Confident) to 100 (Completely Confident).

Accurate Narrative Response on Final Test

	Consistent	Misleading
Exp 2		
Standard	82 (18)	81 (21)
Repeated	92 (13)	87 (23)
Exp 4		
Standard/Full	86 (9)	83 (13)
Repeated/Full	95 (7)	95 (8)
Standard/Divided	85 (14)	87 (15)
Repeated/Divided	93 (15)	93 (15)