Recounting a Memory:
How it Changes Visual Memory and How Useful it is for Others

A Dissertation
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

DEBORAH HUI EN TAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Dr. Yuhong V. Jiang

May 2019

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Yuhong Jiang, for graciously taking me on as her graduate student part way into my third year of graduate school. Your guidance, patience, care, and tough love when I needed it, have been nothing short of instrumental in getting me to where I am today. Your unfailing support for my career development and plans, and words of wisdom as I tried to figure out my career, have also been invaluable. I have learned so much from you and am grateful for the example you have set for me on how to be a great mentor!

I would also like to thank my committee members, Dr. Wilma Koutstaal, Dr. Jonathan Gewirtz, and Dr. Sashank Varma for your guidance on and support for my research. Thank you also to Dr. Roger Remington for providing me with feedback throughout this process during lab meetings.

I would also like to thank the members of the Jiang Lab and Marsolek Lab for supporting me and providing me with feedback on my work throughout my graduate career. Special thanks goes out to all my research assistants who put in countless hours to help me collect and code data for my research! Thank you also to my friends near and far for your endless support. In particular, thank you to my friends from graduate school who understood the power and importance of getting through the hard times together and coming out stronger!

Getting this far in graduate school and in life would not have been possible without the rocks of my life. To my family, thank you for encouraging me, supporting me, and for celebrating my successes. I especially want to thank Dan, my husband, for always being there for me, for lifting me up during the hard times, and for pushing me to never give up.

Lastly, and most importantly, I would like to thank my heavenly Father for setting a path that You know is best for me. I owe every encounter, every decision, the survival of every hardship, and the joy of every success to You.

This dissertation is dedicated to my family.

Abstract

Recounting one's memory is common in everyday experience, such as in educational settings and spatial navigation. How recounting affects memory has largely been examined with verbal materials, with few studies examining its effects on visual memory. The goals of this dissertation are to investigate the effects of recounting a visual memory, first on the person recounting the memory, and second on the ability of the listeners to reconstruct the recounted memory. Following an introduction in Section 1, Section 2 examined how intervening recall affects visual detail memory and whether these effects change across time. Intervening recall interfered with visual detail memory in the short term, but enhanced it in the long term, suggesting that intervening recall has multiple effects on visual detail memory, which change over time. Section 3 showed that memory enhancement from intervening recall can generalize across memory attributes to source memory, supporting the idea that intervening recall can enhance memory as long as fruitful retrieval occurs for the attribute. In addition, there is a positive relationship between what occurs during recounting and subsequent memory performance, raising the possibility that greater fruitful retrieval leads to greater memory enhancement later on. Section 4 examined the usefulness of one's memory recall for listeners. It demonstrated that whereas objects are easy to describe from perception, one's verbal recall of what was seen minutes before is almost useless to third parties. This low utility of one's verbal recall constrains theories about the extension of our memories to the external world and highlights the importance of looking to more informative and reliable sources of information, particularly in higher stakes situations such as during criminal investigations. Overall, this dissertation demonstrates that the way recounting affects

memory depends on many factors. Furthermore, while recounting a memory can benefit the describer, such benefits do not necessarily extend to the listener, suggesting that recounting plays different roles for the describer and the listener.

Table of Contents

List of Tables

List of Figures

Recounting a Memory

1. Introduction

From spatial navigation to eyewitness memory, many activities benefit from an excellent visual memory. This memory includes both the semantic gist of what one saw (e.g., a coffee shop at the corner) and visual details (e.g., a small green shop with a unique design). But how accurate is visual memory? Why are visual memories impressive in some cases (e.g., Brady, Konkle, Alvarez, & Oliva, 2008) and unreliable in others (e.g., Lindholm & Christianson, 1998)? In this dissertation, I explore one factor that may affect the accuracy of visual memory – the act of recounting.

Being asked to recount one's memory is common in everyday experience, such as educational settings, spatial navigation, and crime investigation. Much of past research examining the effect of recounting was situated in the context of education, relying largely on verbal materials. Few have investigated the effects of recounting on visual memory – the focus of my dissertation. Besides affecting one's own memory, recounting also serves the purpose of conveying information to someone else. Whether we are giving directions to another person or describing what a perpetrator looked like, our descriptions guide the listener's future thoughts and actions. It is, therefore, important that useful and accurate information be conveyed. But how good are we at providing that information from memory? In this dissertation, I investigate the consequences of recounting a visual memory for the person recounting as well as the listener.

1.1. Structure of the dissertation

This dissertation has five parts. Section 1 presents a review of the literature on visual memory. This will consist of an evaluation of contradictory findings on the capacity and precision of visual memory, a review of the effects of retrieval practice and of the utility of memory recounting for third parties. Sections 2 to 4 present three studies examining (i) the effects of intervening recall on visual memory, (ii) the effects of recounting on item and source memory, and (iii) the utility of verbal descriptions from memory for others. Section 5 discusses the theoretical and societal implications of these research findings.

1.2. Two views of visual memory

Researchers on visual memory have taken different stances on how good people's visual long-term memory is. Whereas some researchers argue that visual long-term memory has massive capacity for both gist and details, others propose that it is sparse and unreliable. These two traditions of research have typically assessed memory using different procedures (e.g., immediate testing vs. testing after repeated probing), a difference that may have contributed to the different conclusions.

High-capacity visual long-term memory has been demonstrated with photographs of common objects and natural scenes. In one classical study, Standing, Conezio, and Haber (1970) presented participants with 2,560 photographs one at a time. Using a two-alternative-forced-choice (AFC) task immediately following encoding, Standing et al. found that participants could recognize the old photographs with above 90% accuracy.

To test whether the impressive performance was limited to memory for the gist of images, Brady and colleagues performed a series of studies probing memory for specific exemplars using natural scenes or photographs of objects. For instance, Brady et al. (2008) asked participants to encode into memory thousands of photographs of categorically-distinct objects, presented one at a time for 3 seconds each. In a recognition memory test immediately after encoding, participants were highly accurate (above 87%) at distinguishing old pictures from new ones, even when the new objects were from the same basic-level category. Brady and colleagues estimated that visual long-term memory can represent at least 228,000 ($2^{17.8}$) unique codes. Similar findings were observed for scene memory (Konkle, Brady, Alvarez, & Oliva, 2010). High-precision visual memory is not limited to objects to which people currently attend. Hollingworth, Williams, and Henderson (2001) found that people continue to retain in memory previously attended objects, even after their eye gaze (and presumably, attention) has moved away from those objects (see also Hollingworth & Henderson, 2002).

In contrast to studies reviewed above, other studies suggest that visual memory can be surprisingly sparse, even for stimuli that people frequently encounter. In one demonstration, Nickerson and Adams (1979) probed people’s memory for a US penny, an item that participants presumably have had substantial exposure to. Nearly all participants (except for one coin collector) made incorrect drawings of the penny from memory. When presented with drawings of the real penny and its altered versions, fewer than 50% of the participants managed to choose the correct drawing. These findings led to the proposal that visual memory contains primarily semantic or categorical information, such as the unique bronze color of the penny, and lacks details. Consistent with this idea, when participants are tested on their memory for concrete objects (such as drawings of everyday objects) and abstract images, memory is substantially less accurate for abstract images (Koutstaal et al., 2003).

Another line of research suggests that while some complex stimuli yield good visual memory, others do not. Whereas memory for faces of familiar people is accurate, memory for unfamiliar faces is typically poor. In fact, even small changes in lighting condition and viewpoint disrupt recognition of unfamiliar faces (Bruce, Henderson, Newman, & Burton, 2001; Hancock, Bruce, & Burton, 2000). Face memory is poor for unfamiliar faces even when the memory delay interval is as short as a few seconds or hundreds of milliseconds (Jackson & Raymond, 2006; Jackson & Raymond, 2008).

In light of these findings, O'Regan and colleagues proposed that instead of representing visual details in memory, people rely on the external world as their "memory" (O'Regan & Noë, 2001). Visual details are lost as soon as they are no longer in view, and when they are needed for a task, people simply open their eyes and look! This view may seem extreme, but it receives substantial support from research on change detection and visuomotor manipulation tasks. For example, Rensink, O'Regan, and Clark (1997) presented participants with alternating images – an original photograph and an altered version with a portion of the image altered (e.g., deleted, added, or substituted) – and asked participants to spot the difference between the images. The two images were separated by a brief blank interval (of about half a second) and presentation of the images looped back and forth until the change was found. Rensink et al. found that participants needed many repetitions to find the change, especially when the changed information was not of central interest to the portrayed scene. This failure to detect changes has also been found using more naturalistic stimuli, such as the substitution of one person with another person in a video or during real-world interactions (Levin & Simons, 1997). In another line of study, Hayhoe and colleagues measured eye fixation as participants performed a visuomotor task (Hayhoe, Bensinger, & Ballard, 1997). Participants had to pick up several blocks of various colors and place them in another location. Instead of retaining information about multiple blocks in memory, fixation data suggested that participants remembered one piece of information at a time, such as the color of the block that they were about to move. Minimizing the storage of information in visual memory may be advantageous in some ways. Since the external world is readily available to support many tasks, it is cognitively economical to rely on the world as an extended memory space (Rensink et al., 1997).

Together, the studies reviewed above suggest that although people are capable of representing visual details in memory in some situations, they do not typically use or demonstrate that capacity. Additional studies, reviewed next, show that even in situations when it is critical to retain precise information in visual memory, such as eyewitness testimony, memory is often imprecise.

It has long been known that eyewitness memory is unreliable. More than 70% of DNA exonerations in the United States involved eyewitness misidentification, which led to wrongful convictions (“Innocence Project”, n.d.). Laboratory studies show that people have poor source memory, confusing what they saw with what they read (Loftus, Miller, & Burns, 1978). Performance on choosing perpetrators from a lineup is also typically poor (Lindholm & Christianson, 1998). Several reasons contribute to the poor eyewitness memory. First, because memory for unfamiliar faces is poor, an unfamiliar perpetrator will not be easily remembered (Hancock et al., 2000; Patterson & Baddeley, 1977). In addition, stress or arousal (Loftus & Burns, 1982), the complexity of a scene or situation (Liu and Jiang, 2005; Rensink et al., 1997; Levin & Simons, 1997), and the presence of a weapon (Loftus, 1979) may interfere with the encoding and consolidation of visual memory.

Besides the factors indicated above, one factor that may interfere with visual memory in situations such as eyewitness memory is the retelling of one's memory on multiple occasions. Seminal work by Sir Frederic Bartlett in the 1930's (Bartlett, 1932) demonstrates that multiple retellings of one's memory often contain contradictory materials. Expanding on this initial finding, Loftus, Ceci, and others show that memory is highly vulnerable to misinformation provided after encoding but before the final memory assessment (Loftus et al., 1978; Ceci, Loftus, Leichtman, & Bruck, 1994). If retelling one's memory on one occasion introduces misinformation, then this can have a deleterious effect on subsequent recall. Studies that uncover high-capacity visual memory typically test people's memory immediately after encoding, allowing accurate memory to be preserved. In contrast, witness memories are often subject to multiple retellings. How does intervening recall affect visual memory? Would the spectacular memory demonstrated by Brady et al.'s (2008) participants be retained if they had undergone an intervening recall?

In the next section I will review two largely independent literatures related to this question. One line of research on the "testing effect" suggests that retelling or recalling a memory enhances memory and slows down forgetting. The other line of research, on memory reconsolidation, suggests that a memory reminder reactivates previous memory and renders it labile and susceptible to change. This review suggests that intervening recall may have multiple, and sometimes, opposing effects, on memory.

1.3. Divergent effects of memory retrieval

1.3.1. The Testing Effect

The benefit of retrieval practice for long-term retention of information has been known for more than a century (Gates, 1917). A recent wave of new research, championed by Roediger and colleagues, highlighted the beneficial role of this effect on education. For example, Roediger and Karpicke (2006) asked participants to study two prose passages. They gave different groups of participants (i) four opportunities to restudy the material, (ii) three opportunities to restudy the material followed by a free recall test without feedback, or (iii) one restudy opportunity followed by three chances to take the same free recall test. Participants then took a final recall test five minutes or a week later. Results showed that memory was comparable across groups on the initial recall test. In addition, after five minutes, participants who had more restudy opportunities performed better. However, after a longer delay (one week later), participants given additional retrieval opportunities outperformed those who simply restudied the materials. These findings suggest that intervening retrieval improves long-term retention of learned materials. The testing effect has been replicated and extended in later studies using a diverse range of stimuli, including word lists (Carpenter & DeLosh, 2006), word pairs (Liu, Liang, Li, & Reder, 2014; van den Broek, Takashima, Segers, Fernandez, & Verhoeven, 2013; Wing, Marsh, & Cabeza, 2013), general knowledge (Carpenter, Pashier, Wixted, & Vul, 2008), short narratives (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008), foreign language words (Carrier & Pashler, 1992), and face-name pairs (Carpenter & DeLosh, 2005; Landauer & Bjork, 1978).

Several accounts have been proposed to explain why testing is beneficial for long-term retention. These include the Transfer Appropriate Processing account (Morris, Bransford, & Franks, 1977; Bransford, Franks, Morris, & Stein, 1979; Roediger, Gallo, & Geraci, 2002), the Retrieval Effort Account (Carpenter & DeLosh, 2006; Carpenter, 2009) and the Reconsolidation account (Finn & Roediger, 2011; Finn, Roediger, & Rosenzweig, 2012).

The Transfer Appropriate Processing account suggests that how well something is remembered is determined by the extent to which the processing requirements during testing match the processing requirements during encoding (Morris et al., 1977; Roediger et al., 2002). With regard to the testing effect, when participants are given the chance to practice retrieving information from memory, a match of the retrieval practice's processing requirement to that of the final memory test facilitates memory performance. On a broad level, the transfer appropriate processing account is consistent with the testing effect. However, this account also makes inaccurate predictions regarding the match in the testing format between the retrieval practice and the final retrieval. Contrary to what the transfer appropriate processing account would have predicted, several studies show that the best type of retrieval practice is free recall, regardless of whether the final test involves free recall, cued recall, or recognition (Carpenter & DeLosh, 2006). Because free-recall requires the highest retrieval effort, these findings lead to an alternative account: the Retrieval Effort Account.

The Retrieval Effort Account (sometimes also known as the Elaborative Retrieval Account) explains the testing effect in terms of the effort people exert during the retrieval practice. At the core of this account is the idea that successful memory retrieval depends not just on the strength of the memory trace, but also on the accessibility of this trace to retrieval (Bjork, 1975). Accessibility increases as a function of retrieval effort. Because retrieval practice encourages more elaborative retrieval processing than restudy, it allows for more cues to be associated with the material and more retrieval routes to the to-be-retrieved material (Bjork, 1975; Carpenter & DeLosh, 2006). With test practice, the material to be retrieved on the final test can thus be more easily accessed. This account not only predicts the basic testing effect, but also explains why the testing effect is greatest when the retrieval practice involves free recall, rather than cued recall or a multiple choice test.

More recently, Finn and Roediger (2011) suggested that memory reconsolidation contributes to the testing effect. When a memory is retrieved and hence reactivated, it becomes labile and susceptible to change. This memory will then need to be reconsolidated (Dudai, 2004). Finn and Roediger (2011) suggest that once information is retrieved at test practice, memory becomes labile. What occurs right after retrieval leads to the testing effect. In their study, participants studied Swahili-English vocabulary pairs and were later given either a restudy opportunity or a cued recall test. Critically, following restudy or successful retrieval, participants were presented with a blank screen, a neutral picture or a negative picture. These images were shown to differentially affect memory consolidation in the cued recall condition (i.e., negative images were known to facilitate consolidation of recently acquired memory; Anderson, Wais, & Gabrieli, 2006).

In a final test, participants demonstrated the testing effect – memory was better following a cued recall test than following a restudy opportunity. The crucial finding is that the testing effect was greater when negative pictures were used during retrieval practice, supporting the idea that reconsolidation contributes to the testing effect. Finn and Roediger propose that a successful retrieval of the learned information renders that information labile and therefore allows the memory to be strengthened during the re-encoding of the information that had just been successfully retrieved.

The different accounts presented above are not mutually exclusive. Liu and Reder (2016) proposed that the testing benefit stems from both the retrieval and re-encoding processes. In an fMRI study, participants encoded semantically-unrelated English word-pairs and were either given restudy or test opportunities while being scanned. Brain activity during retrieval practice was a better predictor of subsequent memory performance than that during restudy trials. Liu and Reder also examined brain regions that were associated with retrieval and those associated with re-encoding after retrieval. Brain activity associated with retrieval consistently predicted subsequent correct recall, whereas brain activity associated with re-encoding did not predict subsequent correct recall if participants had already successfully recalled the word multiple times. These results suggest that the testing effect involves both retrieval of previous memory and re-encoding of the information to memory.

Each account of the testing effect presents an explanation for why retrieval practice can benefit long-term retention. While these benefits of testing on memory are undeniable, it is important to note situations when retrieval can be detrimental instead.

The memory reconsolidation literature, which highlights such situations, will be discussed next.

1.3.2. Interference through reconsolidation

Retrieval of a memory, while at times beneficial, can also interfere with the original memory through reactivation and reconsolidation. Specifically, when a reminder of a memory (e.g., through the presentation of a cue) is presented, it reactivates the memory as it is retrieved. Upon reactivation, the memory trace becomes labile and susceptible to change for a certain period of time, typically less than six hours (Schiller et al., 2010; Duvarci & Nader, 2004; Walker, Brakefield, Hobson, & Stickgold, 2003). Through protein synthesis, memory is reconsolidated (Nader, Schafe, & LeDoux, 2000; Suzuki et al., 2004). New information introduced during this period may alter the previous memory trace, producing interference (Dudai, 2004; Diekelmann, Büchel, Born, & Rasch, 2011; Schiller et al., 2010).

Memory interference via reconsolidation is evident in non-human animals such as rodents (Nader et al., 2000), chicks (Summers, Crowe, & Ng, 1997; Litvin & Anokhin, 2000), and fish (Eisenberg, Kobilo, Berman, & Dudai, 2003). For example, Nader et al. (2000) paired a tone with foot-shock. Through conditioning, rats exhibited a fear response when presented with the tone. Twenty-four hours after the establishment of this memory, the rats were presented with the tone alone as a reminder cue. To disrupt protein synthesis, Nader et al. infused anisomycin, a protein synthesis inhibitor, into the amygdala immediately after the reminder cue or after several hours of delay. Anisomycin infusion immediately with the reminder cue (but not after a delay) disrupted the learned fear memory, demonstrated by a lack of freezing response in subsequent testing sessions. These results suggest that when memory is reactivated, it is susceptible to disruption by protein synthesis inhibitors. Reactivation and reconsolidation are also seen in spatial learning in rodents (Suzuki et al., 2004).

Memory interference via reconsolidation has also been demonstrated in humans, even without the use of protein synthesis inhibitors (Schiller et al., 2010; St. Jacques, Olm, & Schacter, 2013; Kindt, Soeter, & Vervliet, 2009). Schiller et al. paired a shock to the wrist with a specific colored square, establishing conditioned fear in participants. One day later, all participants underwent extinction training whereby the colored square was presented repeatedly without the shock. However, for some participants, prior to extinction training, the colored square was presented once as a reminder either 10 minutes or 6 hours before extinction training. Note that 10 minutes falls within, but 6 hours falls outside of, the memory reconsolidation window. Fear memory was probed one day later. Participants who received the reminder 10 minutes before extinction training showed diminished fear response compared with participants who received no reminder, or who received the reminder 6 hours before the extinction training. These effects lasted at least a year. Thus, fear memory can be updated with non-fearful information presented when the memory is reactivated.

Studies reviewed above raise the question of how misinformation introduced during an intervening memory recall affects the original memory trace and hence, subsequent memory performance. Misinformation is defined as misleading information (Loftus, 2005) that, when presented to a person after initial encoding, impairs subsequent memory performance (i.e., the misinformation effect; Loftus & Palmer, 1974). Though this effect is often interpreted in terms of source memory error, it could also reflect reconsolidation of the original memory when it is reactivated by the misinformation. Consistent with this possibility, Chan, Thomas, and Bulevich (2009) demonstrated that when a witnessed event is recalled, it is more susceptible to interference from misinformation introduced afterwards. They suggested that recalling the memory reactivated the memory, causing it to become labile and susceptible to change. Once memory was reconsolidated, the misinformation became incorporated into the original memory and impaired subsequent memory performance.

Misinformation studies are often conducted by having the researchers introduce the misinformation (e.g., Chan et al., 2009; Loftus & Palmer, 1974; Loftus & Pickrell, 1995; Belli, Lindsay, Gales, & McCarthy, 1994). On the other hand, the misinformation can be introduced by the participants themselves. For example, studies on children's memory suggest that misinformation can be spontaneously introduced by children themselves during memory retrieval (Brainerd & Reyna, 1998a; Reyna, 1995). In adults, describing a face they saw earlier led to worse memory of the described face relative to the not-described face. Though this finding is sometimes interpreted in terms of altered coding (i.e., from visual to verbal codes) of the original memory (Schooler & Engstler-Schooler, 1990, "verbal overshadowing"), it may also arise because the verbal description contains misinformation that gets incorporated into the original memory. Support for this latter explanation can be gleaned from a recent large-scale replication, involving 31 labs, of the verbal overshadowing effect (Alogna et al., 2014). The replication showed that memory impairment from verbally recalling the perpetrator's face was larger when there was a delay after an event was witnessed and before a verbal description was provided, compared to when the verbal description of the perpetrator was provided immediately after witnessing the event. Given that memory may have weakened during the delay, it raises concerns about whether spontaneous misinformation occurs, leading to memory interference later on.

In summary, retrieving one's memory before a final memory test may have opposing effects. Retrieval can enhance subsequent memory through the testing effect, but it may interfere with memory via reactivation and reconsolidation. With the exception of a few studies on faces and other pictures, these findings have primarily been observed with verbal materials (such as prose passages or foreign vocabulary learning) or in tests of fear memory. How recalling one's visual experience affects visual memory is less understood.

1.3.3. Retrieval of visual memory

In this section, I review studies that examined effects of retrieval on visual (as opposed to other forms of) memory.

1.3.3.1. Testing effect on visual memory

Studies that used visual stimuli have found evidence that intervening retrieval can benefit subsequent memory performance. Carpenter and Pashler (2007) asked participants to study two maps for a memory test. One map was restudied, the other was studied once and then given retrieval practice. For retrieval practice, participants were shown the map with a feature omitted and had to recall the missing feature and its location. In the final test, participants drew the maps from memory. The participants were better at reproducing the map that received intervening testing than the one that was restudied.

Even memory for basic visual features such as colors benefits from the testing effect. Sutterer and Awh (2016) presented participants with 400 unique shapes of real objects such as a dinosaur and a kiwi bird. The pictures were colored in one solid color that could be found on a color wheel. After the presentation of every 10 pictures, participants were (i) given a retrieval practice where they were shown a white version of the object and had to choose the color on the color wheel that the object was previously presented in, (ii) a control condition that involved no further presentations of the objects, or (iii) a restudy opportunity in which they could view all 10 objects with their corresponding colors again. At final testing, participants recalled the color of a higher proportion of the objects that they were previously tested on, relative to the other conditions. This finding suggests that the testing effect increased the accessibility of color memory.

Despite these findings, retrieval practice has been shown to incur a cost to memory in some cases. Such a cost is commonly demonstrated with the Retrieval-Induced Forgetting (RIF) effect (Anderson, Bjork, & Bjork, 2000; Ciranni & Shimamura, 1999). Specifically, recalling information can impair memory for related but unrecalled information. Ciranni and Shimamura showed participants 4 circles, 4 triangles, and 4 crosses that each had a distinct color and location on a wheel. After initial study, participants were given retrieval practice where they were asked to recall one of the features (e.g., color) of a shape for a subset of the stimuli (e.g., 2 circles) with the non-tested features as cues (e.g., shape, location). In a final test they had to recall the previously asked about feature (color) of all the stimuli. Ciranni and Shimamura found that memory was better for the items that received retrieval practice, but was impaired for items related to them (i.e., the other 2 circles that were not tested in the intermediate phase). When the subset of stimuli was given restudy opportunities instead, no impairment of related items was observed at final testing. This study suggests that retrieval practice can facilitate retention of memory for individual visual stimuli, but incur a cost to memory for related but untested items.

In summary, retrieval practice can benefit subsequent visual memory. However, the intervening test may incur a cost to related but untested items.

1.3.3.2. Reconsolidation effects on visual memory

As reviewed in Section 1.3.2., reactivation and reconsolidation can interfere with memory. This finding extends to visual materials. Specifically, studies on rodents have found that object recognition memory is subject to reconsolidation (Akirav & Maroun, 2006; Rossato et al., 2007). For example, microinfusion of anisomycin (a protein inhibitor) into the ventromedial prefrontal cortex immediately after the reactivation of object memory impairs object memory (Akirav & Maroun, 2006). Infusion of anisomycin into the hippocampus in rats also interferes with object memory (Rossato et al., 2007). In people, recognition memory for content in a TV show that they had seen was impaired by misinformation introduced by the experimenter only if the memories had been reactivated beforehand (Chan & LaPaglia, 2013). Similar results were shown in memory for what was seen on a museum tour (St. Jacques et al., 2013). One day after going on a museum tour, some participants' memories were reactivated by seeing pictures that they had taken on the tour. After which, they were presented with false images that they had not seen previously on their tours. On the final day of testing, participants whose memories were reactivated before being exposed to the false images had higher false alarm rates for the false images, compared to those whose memories were not reactivated before the false image presentation. These results suggest that new information can be incorporated into the original visual memory trace after its reactivation. Having said this, the number of human studies that have examined interference effects on visual memory from reconsolidation are still few in number (Elsey, Van Ast, & Kindt, 2018). Thus, it leaves open the question of whether these interference effects can be generalized across other types of visual stimuli.

1.3.4. Retrieval effects across memory attributes

Besides examining the generalizability of recounting effects across domains (verbal and visual), one could also extend this examination to different aspects of a particular memory. Past work showing enhancement from intervening testing on visual memory has mainly demonstrated a facilitation of visual *item* memory (e.g., individual objects). Items in the world rarely occur in isolation, however. They typically appear in the context of something else, like the environment in which it is seen. This raises the question of whether intervening retrieval affects source memory in a similar way to that found for item memory.

Past research has found evidence that if source information is retrieved during an intervening retrieval test, a testing effect can be observed for source memory later on (Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010). This research did not examine item memory, making it unclear whether intervening retrieval has comparable effects across attributes of a memory. The retrieval effort account (Carpenter & DeLosh, 2006; Carpenter, 2009) proposes that effort exerted to retrieve a memory enhances the memory. By this logic, the effects of intervening retrieval should apply similarly to both item and source memory as long as effort is exerted to retrieve both types of information. Whether this is indeed the case remains to be tested.

Recounting a memory could, therefore, have multiple effects on memory, and these effects may be generalizable both across material (e.g., verbal and visual) and across different attributes of the memory trace (e.g., item and source information). Sections 2 and 3 of this dissertation will examine the effects of recounting a visual memory on one's memory.

Though most of the existing research has focused on what recounting does to one's own memory, in our daily lives we often recount a visual memory as part of social communication. But how well can people convey information about what they had seen? To what degree can we rely on someone else's memory recall? As reviewed next, addressing this question has both theoretical and societal implications.

1.4. Usefulness of one's memory recounting for third parties

In our daily lives, an important function of recounting one's memory is to provide information to other people. Providing descriptions of what one had seen is common when giving directions, reporting a crime, and telling a story. But how well can we convey information in visual memory to others? Studies that have examined this question are few, and those that have done so have found seemingly contradictory results.

Work that analyzed eyewitness descriptions from police records suggests that witnesses typically provide accurate but sparse information in their descriptions (Van Koppen & Lochun, 1997). Witnesses tend to give accurate, but general descriptions such as the perpetrator's gender, race, and build, while leaving out more specific information, such as the perpetrator's facial characteristics. Thus, although accurate information may be provided, its sparseness is of little use to law enforcement who use it to search for the perpetrator.

Laboratory studies have examined the utility of one's recounting of a visual memory to others by measuring how well third parties can identify the perpetrator based on the description a witness provides. In one study (Christie & Ellis, 1981), researchers showed participants a target face for 60 seconds, and afterwards had them write down a description of the target face in as much detail as possible. If their descriptions missed out any of five key features (forehead/hairstyle, eyes, nose, mouth, and chin), the participants were prompted to provide information about the missing features in their descriptions. A group of judges was then given the participants' descriptions and asked to select the target face from an array of faces. All in all, judges were able to identify about half of the target faces, suggesting that when witnesses are allowed to look at the visual stimuli for a significant period of time, and are probed immediately after with specific prompts, their verbal descriptions can be useful for third parties.

In general, previous research directly examining the utility of verbal descriptions are few, and those that have done so, have mainly focused on faces as stimuli. Faces, especially unfamiliar ones, are notoriously difficult to remember (Bruce et al., 2001; Hancock et al., 2000). Much of facial recognition also depends on recognizing the configuration of the features in the face (Young, Hellawell, & Hay, 1987), a type of information that is hard to verbalize. This raises the question of whether the usefulness of verbal descriptions is enhanced for visual stimuli that are more memorable and easily describable, such as photographs of common objects.

1.5. Summary

The contrasting outcomes, and often limited focus of past studies that have investigated how recounting a memory can change subsequent memory, highlight the importance of examining this question in order to better understand memory. Studies that have found testing benefits from intervening retrieval, for example, have primarily been done in the context of education and thus used verbal stimuli, limiting the generalizability of the testing benefit to other domains. Although there have been studies that examined how intervening recounting affects visual memory, they are few and are limited in several ways. These studies do not separate memory for visual details from memory for basic categories. Studies that have found benefits from recounting evaluated memory at a coarser level (e.g., Carpenter & Pashler, 2007) compared to other studies that have found interference from recounting (Schooler & Engstler-Schooler, 1990). Carpenter and Pashler (2007), for instance, recorded position accuracy of the locations of objects on the maps in terms of whether the object was placed in the correct quadrant of the map and in the correct position relative to other features on the map. Schooler and Engstler-Schooler (1990), on the other hand, asked participants to select the target face from an array of similar faces, essentially testing more detailed memory. Without specifically separating exemplar (i.e., detail) memory from category memory, it leaves open the question of whether intervening retrieval affects memory for visual details in a different way than that for category information. This is the primary research question that I address in Section 2. In addition to this question, the variety of temporal delays between intervening retrieval and final memory test in past studies make it difficult to compare across studies and determine whether the enhanced benefit observed with a delay after recounting in studies using verbal material also extend to studies using visual material. I also address this question in Section 2.

Assessing the generalizability of the effects of intervening retrieval additionally involves questions about whether recounting similarly affects different attributes of memory as well. As discussed in the literature review above, past studies have either examined the effects of recounting on item or source memory (e.g., Brewer et al., 2010). Without examining the effects on both item and source memory in a single study, it is difficult to know whether intervening retrieval has similar effects across memory attributes. While the leading account of the testing effect – the retrieval effort account – seems to suggest that item and source memory should be affected similarly as long as retrieval effort is exerted for both types of information during recounting, this prediction has not yet been tested. This is the question I will address in Section 3.

The act of recounting is done more often for the purpose of conveying information to someone else, rather than to be used as a mnemonic device for oneself. The dearth of past research that has examined the utility of verbal descriptions for others makes it unclear exactly how well one can convey information about what was seen to someone else. In particular, using stimuli that are more easily described and more memorable than faces (a common type of stimuli used in past studies), are people able to produce useful descriptions for third parties so that the latter can identify what the former had seen? I address this question in Section 4 of my dissertation.

Besides examining the usefulness of one's verbal descriptions for others, there is also the question of whether one's verbal recall has any correspondence to one's subsequent recognition accuracy. This question has implications for a U.S. Supreme Court guideline (Neil v. Biggers, 1972) suggesting that a more accurate prior description by a witness should imply that their subsequent identification is also more reliable. Past findings, however, show a poor correspondence between recall and recognition, raising doubts about the accuracy of the U.S. Supreme Court's guideline. On that note, in Section 4, I investigate whether the usefulness, and therefore accuracy, of one's recall has any correspondence to one's ability to recognize the visual stimulus later on. A direct relationship between participants' verbal recall and their subsequent recognition accuracy will lend support to the argument that verbal recall and visual recognition rely on similar sources of memory.

2. Section 2

Recounting a visual memory is common in daily life. Past studies have shown that recounting a memory can facilitate subsequent memory, yet most studies focused on verbal materials or conflated category memory with memory for visual details, leaving open the question of whether visually-specific interference effects on visual detail memory may occur with recounting. Section 2 presents a set of experiments that answer this question. These experiments investigate how intervening retrieval affects detail memory and how the effect changes across time. This design allowed for the testing of competing theories, such as the testing effect and interference accounts. Using this procedure, we showed that intervening recall has multiple effects on visual memory, which change in magnitude across time.

Section 2 was published in 2018: Tan, D. H., & Jiang, Y. V. (2018). How visual memory changes with intervening recall. *Quarterly Journal of Experimental Psychology* [Epub Ahead of Print]. https://doi.org/10.1177/1747021818808524. The article is reproduced here in its entirety.

How visual memory changes with intervening recall

Deborah H. Tan

Yuhong V. Jiang

Department of Psychology, University of Minnesota

Abstract

Being asked to recount a visual memory is common in educational settings, spatial navigation, and crime investigation. Previous studies show that recounting one's memory can benefit subsequent memory, but most of this work either used verbal materials or conflated category memory with memory for visual details. To test whether recounting may introduce visually-specific interference effects, we tested people's memory for photographs of objects, but introduced an intervening phase in which people described their memory. We separated memory for the specific exemplar from memory for the basic-level category. Contrary to recent findings on maps and colours, the intervening retrieval practice did not consistently strengthen exemplar memory of objects. Instead, recounting one's visual memory appeared to introduce interference that sometimes cancelled the benefit of increased retrieval effort. Delaying the final memory test by 24 hours increased the benefit of retrieval practice. These findings suggest that intervening retrieval has multiple effects on visual memory. Instead of being a snapshot, this memory constantly changes with retrieval practice and with time.

Keywords: visual memory, intervening recall, testing effect, verbal overshadowing

Introduction

Memory is traditionally assessed in two stages. First, people encode pertinent information to memory. Next, their memory is tested following a delay. This two-stage procedure overlooks the fact that people frequently recount their memory between encoding and the final test, a process that may alter memory. The past decade has seen a surge of research on effects of such intervening recall. Entitled “the testing effect,” this new line of work demonstrates dramatic enhancement of memory by intervening retrieval practice (Roediger & Karpicke, 2006). The benefit stems largely from increased effort people exert during retrieval practice, allowing more cues to be associated with the encoded material and enhancing its accessibility during final testing (Bjork, 1975; Carpenter & DeLosh, 2006). The positive testing effect is sometimes accompanied by retrieval-induced forgetting, in which intervening retrieval practice impairs memory for related but not practiced stimuli (Anderson, Bjork, & Bjork, 2000). Because much of this work is conducted in the context of education, it relies heavily on verbal materials such as word lists or passages. Few studies have directly examined effects of intervening recall on visual memory, the focus of the current study.

Two reasons suggest that the testing effect may be similar between verbal and visual memory. First, the leading theory of the testing effect – the retrieval effort account – makes no reference to the type of material. Effects of retrieval effort are to increase the accessibility of the encoded material, as such, they should apply to both verbal and visual stimuli. Second, visual memory typically contains both visual details of the image and its basic-level category. The latter is not strictly visual as people tend to name visual objects at their basic-level category (Rosch, 1975). This basic-level category, thus, may share a representational format with verbal memory, raising the likelihood that information about object category is coded both visually and verbally. As long as basic-level category is a part of the memory tested, similar effects are expected for visual and verbal materials.

Consistent with the material-independent account just mentioned, recent studies on visual memory have revealed positive effects of intervening retrieval. Carpenter and Pashler (2007) asked participants to study two maps. Subsequently, one map was studied once again (restudy). For the other map, participants viewed an altered version and recalled missing features. Immediately afterwards participants drew the maps from memory. Reproduction was better for the map that received retrieval practice than for the one restudied. Extending these findings to colours, Sutterer and Awh (2016) presented participants with coloured shapes such as a blue dinosaur or an orange bird. Participants saw some of the coloured shapes for a second time, and viewed a white version of the other shapes and recalled their colour. In a final test administered either immediately or more than a day later, participants recalled the colour of more shapes from the retrieval practice condition than from the restudy condition. Finally, Maxcey and Woodman (2014) examined retrieval-induced-forgetting using photographs of common objects. They presented two exemplars per category (e.g., two sports cars, two muffins) during the initial encoding phase, then introduced retrieval of one exemplar (e.g., sports car A), and immediately afterwards tested old-new memory for both exemplars as well as objects not included in retrieval (e.g., the muffins). Results showed two effects of intervening retrieval: it enhanced memory for the exemplar included in retrieval (e.g., sports car A), but impaired memory for the encoded but not practiced exemplar (e.g., sports car B), relative to items from categories not included in retrieval. These findings suggest that recounting one's visual memory strengthens that memory, though it may induce forgetting of related but not-practiced stimuli.

An important difference between visual and verbal stimuli, however, suggests that intervening retrieval may have unique effects on visual memory. The visual component of the memory does not easily lend itself to a verbal code. Neuroimaging studies show that visual cortical regions are recruited to retain visual memory. For example, remembering faces preferentially activates the fusiform face area whereas remembering scenes preferentially activates the parahippocampal place area (Ranganath, DeGutis, & D'Esposito, 2004). If the intervening retrieval involves verbalizing a visual memory, such as when one recounts a complex graph learned in class or a recent hiking trip, the introduction of a verbal code may distort the original visual memory.

Supporting this interference account is the classic demonstration of verbal overshadowing. Schooler and Engstler-Schooler (1990) showed participants a videotape of a crime scene and asked some participants to describe the perpetrator's face. Five minutes later, all participants performed a final recognition test of the perpetrator's face. Those who previously described the perpetrator performed worse than those who did not verbalize the description in selecting the perpetrator from a lineup. Schooler and Engstler-Schooler suggest that having verbally described the faces, participants may have re-coded their memory in verbal codes, interfering with visual memory. Verbal overshadowing suggests that retrieval practice can interfere with the original visual memory. However, the replicability of this effect is controversial (Yu & Geiselman, 1993), raising questions about its generality.

Interference effects may extend beyond verbal overshadowing. Even when the intervening retrieval does not involve verbal re-coding, this act may interfere with visual memory if inaccurate or misleading visual details are retrieved. When an experimenter introduces misleading information during a retrieval attempt, this information is incorporated and re-consolidated to form new, and incorrect memory (Chan, Thomas, & Bulevich, 2009). Misinformation effects are widespread in eyewitness memory (Loftus, 2005). They may also happen when participants themselves, rather than the experimenters, introduce misinformation during a retrieval attempt (Brainerd & Reyna, 1998a; Reyna, 1995). The potential presence of interference may explain why, in Sutterer and Awh (2016), retrieval practice increased the number of objects recalled, but not the precision of the recalled memory. It may have also contributed to the retrieval-induced-forgetting observed in Maxcey and Woodman (2014).

The above review suggests that intervening recall may have multiple effects on visual memory, particularly memory for visual details. However, several factors obscure an understanding of these effects. First, although previous studies tested visual memory, they did not adequately separate memory for visual details from memory for basic categories. Map memory in Carpenter and Pashler (2007) was evaluated at a coarse level, in terms of whether certain features were drawn in the correct relative spatial locations. Colours used in Sutterer and Awh (2016) could be named, and given the lack of an enhancement in memory precision, the enhanced memory accessibility may be attributed to the categorical rather than visual component. Maxcey and Woodman (2014) used an old-new task, where performance may be supported by both category and detailed memory. The lineup task used in Schooler and Engstler-Schooler (1990) is the only one that presents exemplar-level foils at testing. As reviewed above, this line of work shows interference. However, memory for unfamiliar faces is notoriously poor (Bruce, Henderson, Newman, & Burton, 2001; Hancock, Bruce, & Burton, 2000), limiting the generality of the latter finding. One important goal of the current study is to isolate memory for visual details by using a forced-choice recognition task that includes exemplar-level foils.

A second limitation of the previous studies is the variety of temporal delay between intervening retrieval and the final memory test. Most common is an immediate test (Carpenter & Pashler, 2007; Maxcey & Woodman, 2014; Sutterer & Awh, 2016). Some administered the test more than one-day later (Sutterer & Awh, 2016's Experiment 3). This factor makes it difficult to compare across studies. In the larger literature on the testing effect, an increase in the delay interval is known to strengthen the testing effect. In fact, Roediger and Karpicke (2006)'s classic study found a testing effect using passages only after a delay (2 days or 1 week later). Whether analogous effects are seen with visual stimuli are unclear. An examination of delay interval is theoretically and practically important. Theoretically, it allows us to uncover effects that may be apparent only after a delay. Practically, this research has implications for memory tasks in applied situations (e.g., classroom learning; eyewitness testimony), which often extend over long delay intervals. The second goal of the present study is to investigate effects of delay interval by administering both immediate and 24hr-delayed tests.

Experiment 1

This experiment establishes the basic experimental procedure. Stimuli were photographs of real objects (Brady, Konkle, Alvarez, & Oliva, 2008; Maxcey & Woodman, 2014). These stimuli not only yield rich visual memory, but also permit a straightforward separation of category and exemplar memory. The procedure consists of three phases: encoding, intervening cued recall, and recognition. In Phase 1 (encoding), photographs of objects were shown one at a time for encoding. All objects were semantically distinct. For example, the encoding phase could not contain two different apples; this differs from Maxcey and Woodman (2014) who presented two exemplars per category for encoding. Next, in Phase 2 (intervening recall), participants typed out a description for a subset of the objects. Finally, in Phase 3 (final test), we assessed recognition memory for all objects. To isolate exemplar memory from memory for basic category, each trial of the final test contained four images: the old object, a different exemplar from the same basic-level category as the old object, and two exemplars from a new basic-level category. All phases were administered on the same day in Experiment 1; a later experiment examined longer delays between Phases 2 and 3.

With the design above, we computed two types of visual memory. Category memory was indexed by the proportion of times that participants chose the correct category – either the old object or the exemplar foil, rather than the new category. Exemplar (detailed) memory was indexed by the proportion of times that participants chose the old object rather than the exemplar foil. In probability terms, exemplar memory was computed as P(old object | correct category). Our research focus is primarily on exemplar memory, a type of memory likely distinct from verbal memory. We examined whether intervening cued recall enhanced this memory, as expected from recent studies on the testing effect, or whether it interfered with exemplar memory, as expected from verbal re-coding or misinformation effects.

Method

Participants. Participants in this study were students from the University of Minnesota, all Native English speakers, naïve to the purpose of the study, had normal or corrected-to-normal visual acuity, normal colour vision, and no history of neurological or psychiatric conditions. No participant was tested in more than one experiment. Study protocol was approved by the University of Minnesota IRB.

Sample size. A power analysis was conducted, using the power.t.test function in R, based on Experiment 1 of Sutterer and Awh (2016), with a reported Cohen's d of 3.64 (p. 835, $t(11) = -6.03$). A sample size of 6 achieves a power greater than 0.80. In addition, to counterbalance experimental materials across conditions in this study, sample size needs to be in multiples of 12. Based on these considerations, each of Experiments 1A-C tested 12 participants, and subsequent experiments that examined additional factors tested 24 participants. The final sample size fluctuated slightly due to participant signups.

Thirty-seven participants completed Experiment 1. There were 13 in Experiment 1A (10 females and 3 males, $M = 19$ years old, $SD = 1.1$), 12 in Experiment 1B (8 females and 4 males, $M = 19.5$ years old, $SD = 1.3$), and 12 in Experiment 1C (10 females and 2 males, $M = 19.5$ years old, $SD = 2.0$).

Equipment. Participants were tested individually in a room with normal interior lighting. They sat approximately 40cm away from a 19" CRT monitor (resolution 1024x768 pixels; refresh rate 75Hz). The experiment was programmed using Psychtoolbox (Brainard, 1997; Pelli, 1997) implemented in MATLAB (www.mathworks.com).

Materials. Photographs of objects were sampled from Brady and Oliva's object database (http://cvcl.mit.edu/MM/stimuli.html). There were 2 exemplars from 310 distinct object categories.

Procedure and Design. Participants were tested in three phases, separated by 2 minutes between phases (Figure S2.1.).

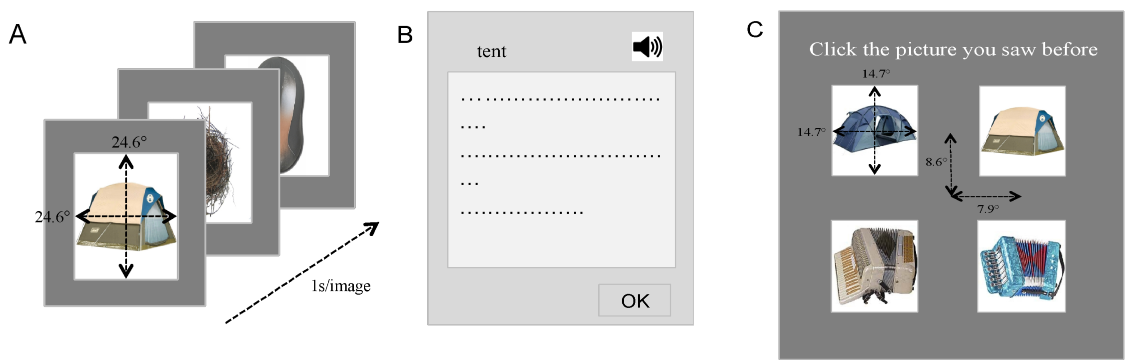

Figure S2.1. A schematic illustration of the encoding, cued recall, and recognition phases. A. The encoding phase presented each object at a pace of 1s/image. All images came from different basic-level categories. B. An example of the intervening cued recall phase. Participants were asked to type out a description of the named object that they saw in the preceding phase. C. The recognition phase with the old object, an exemplar foil, and two between-category foils not seen previously. Information about the dimension of stimuli was not actually shown on the display.

(1) Phase 1: encoding. In the first phase, participants viewed a series of photographs presented at the center of the screen (24.6°x24.6°), at a pace of 1s/image. They were asked to remember the objects. To ensure that they engaged in the task, we asked them to monitor occasional image repetitions. The sequence contained 170 images,

including 150 objects that appeared once, and 10 fillers that appeared twice. The 170 images were presented in a random order. The fillers were included for the repetition detection task and were not used in subsequent phases. The same 10 categories served as fillers for all participants, leaving us with 300 categories from the image set. A random half of the 300 categories was assigned for encoding and the other half as foils in the testing phase (phase 3). This assignment was reversed for half of the participants. The 150 target objects were categorically distinct at the basic level. Half of the participants encoded one of the exemplars of a given category, while the other exemplar served as an exemplar foil in the testing phase. This was reversed for the other half of the participants. Object categories and exemplars used for targets and foils were fully counterbalanced across participants.

The three versions of the experiment differed in minor ways. In Experiments 1A and 1B, photographs were presented visually only. Participants pressed a button whenever they detected a repetition. In Experiment 1C, a voice naming the object accompanied the photograph. The voice was added to ensure a clear correspondence between the verbal cues used in Phase 2 and the objects. The audio files were edited to last no longer than 1s. The audio tied up the computer processor, preventing online keyboard responses. Participants therefore kept a count of the number of repetitions and reported that number at the end of Phase 1.

(2) Phase 2: intervening cued recall. A random third of the encoded objects (N=50) were used in the second phase. The assignment of stimuli for inclusion in Phase 2 was counterbalanced across participants. In this phase, participants were shown the written name of an object on the screen. A voice announcing the object's name was played

alongside in Experiment 1C. Participants were asked to type out a detailed description of the named object they saw in Phase 1. After they finished typing they clicked "OK" to proceed to the next one. No feedback was given. To further increase recall effort, Experiments 1B and 1C included an additional instruction for participants to describe the objects in such a way that someone else could pick out the objects based on their description.

(3) Phase 3: final memory test. The final memory test included 150 trials. On each trial, four objects were presented, one in each quadrant (image size: 14.7ºx14.7º; the center of each image was 11.7º from fixation, horizontally displaced by 7.9° and vertically displaced by 8.6° from fixation). One object, the old object, was previously presented in the encoding phase. Another object was from the same basic-level category as the old object, but was a new exemplar (the "exemplar foil"). The other two objects were two exemplars from a basic-level category not presented previously (the "between-category foils"). The position of the four objects were random. Participants were asked to click on the object they saw before. A tone provided feedback.

Results

1. Encoding phase

Participants conducted a repetition detection task in Phase 1. For Experiments 1A and 1B where participants pressed a button whenever a repetition was detected, mean hit rate was 77% (S.E. = 3%) and mean false alarm rate was 2% (S.E. = 0.5%). For Experiment 1C where participants tallied the total number of repetitions, mean repetitions detected was 7.3 (S.E. = 2) out of 10. Repetition detection rates were also high in subsequent experiments. Participants reported 8.5 repetitions in Experiment 2, 10 in Experiment 3, 10 in Experiment 4A, and 9.4 in Experiment 4B. These results verified that participants were engaged in the encoding task.

2. Intervening recall

The verbal descriptions participants provided frequently contained a mixture of accurate and incorrect descriptions. To assess the accuracy of the description, we asked three coders to identify the object participants described. The coders were shown participants' descriptions along with the old object and the exemplar foil. Coders accurately identified just 57% of the old object (S.E. = 1%), a level close to but significantly higher than chance, $t(36) = 47.75$, $p < .001$. Similar low levels of coder accuracy were seen in Experiment 2 (59%), Experiment 3 (57%), Experiment 4A (61%), and Experiment 4B (59%). Thus, verbal descriptions participants gave were not useful in guiding a third person. However, the results do not mean that participants' own memory was poor – the description was given without knowing the nature of the exemplar foils, so participants could not have anticipated key features necessary for the coders' task. In fact, coder accuracy was considerably lower than participants own recognition rates (see Final memory test section).

When analyzing coders' performance based on participants' recognition results in the third phase, coders were significantly above chance (63%) for images that participants themselves successfully identified later on, $t(36) = 9.05$, $p < .001$. The coders' accuracy, however, declined to a level significantly below chance (39%) for objects to which participants later chose the exemplar foil instead of the target, $t(36) = 6.44$, $p < .001$. This suggests that when participants had poorer exemplar recognition memory, their earlier verbal description of the objects contained misinformation, misguiding coders to choose the exemplar foil more often than the target. Similar results were obtained in subsequent experiments and will not be further reported.

3. Final memory test

On each trial of the final memory test, participants were shown an old object, a new exemplar from the same category as the old object (“exemplar foil”), or two objects from a new category (“category foils”). Responses could fall into any of these three types (Table S2.1.). Two of the responses reflected correct category memory. Hence category memory was computed as the proportion of trials that participants chose the correct category (the old object or the exemplar foil) out of all choices. Exemplar memory was computed among trials that participants chose the correct category. It was the proportion of trials that participants chose the old object out of trials where a correct category was chosen.

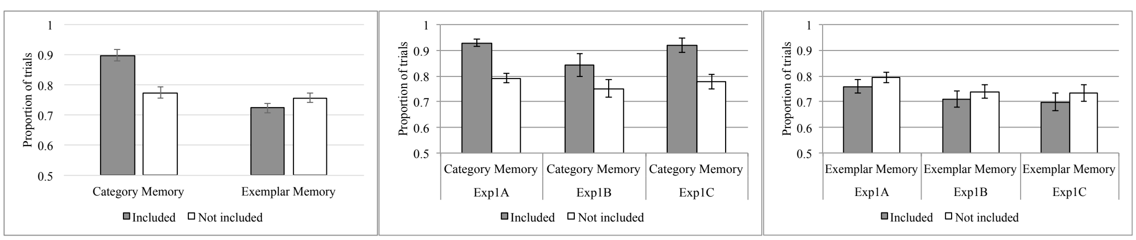

Figure S2.2. Recognition memory results from Experiment 1 for objects included and not included in the intervening recall phase. Category memory was the proportion of trials where the correct category was chosen (either the old object or the exemplar foil) out of all trials. Exemplar memory was the proportion of trials where the old object was chosen out of trials where the correct category was selected. Left: Results from all three versions. Middle: Category memory results from Experiments 1A, 1B, and 1C. Right: Exemplar memory results from Experiments 1A, 1B, and 1C. Error bars show ±1S.E. of the mean.

As shown in Figure S2.2., intervening cued recall had opposite effects on category memory and exemplar memory. An ANOVA on type of memory (category or exemplar), intervening recall (included or not included), and experimental version (1A, 1B, or 1C) revealed a significant interaction between type of memory and intervening recall, $F(1, 34) = 57.37$, $p < .001$, $\eta_p^2 = .63$; there was no three-way interaction, $F(2, 34) = .77$, $p > .40$.

We next separately analyzed category and exemplar memories. For category memory, an ANOVA on intervening recall and experimental version revealed a significant main effect of recall status, $F(1, 34) = 107.77$, $p < .001$, $\eta_p^2 = .76$, an effect that did not interact with experimental version, $F(2, 34) = 1.86$, $p > .15$. Performance was 12% higher for objects included in Phase 2 than for those not included. In contrast, exemplar memory showed the opposite. The main effect of intervening recall, $F(1, 34) = 5.17$, $p < .02$, $\eta_p^2 = .13$, which did not interact with experimental version, $F < 1$, was driven by a 3% reduction in performance for objects included in the intervening recall.

Table S2.1. Mean percentage of trials where participants chose the old object, the exemplar foil, or one of the category foils during the recognition phase (phase 3), for objects included and not included in the preceding intervening cued recall phase (phase 2). Standard error of the mean is shown in parenthesis.

Experiment	Recall Condition	Old	Exemplar foil	Category foil
Exp 1A	Not in recall	63.2 (3)	16 (1.4)	20.9 (2)
	Included in recall	70.6 (3.1)	22.3 (2.4)	7.1 (1.4)
Exp 1B	Not in recall	56.4 (4.3)	18.7 (1.3)	24.9 (3.3)
	Included in recall	60.5 (5.1)	23.7 (2.6)	15.8 (4.5)
Exp 1C	Not in recall	57.9 (4.4)	19.9 (1.8)	22.2 (2.9)
	Included in recall	65 (4.5)	27 (2.6)	8 (2.8)
Exp 2	Not in recall	60.8 (2.4)	18.8 (1.3)	20.4 (2)
	Visualized + Described	66.7 (2.3)	25.2 (1.6)	8.1 (1.3)
	Visualized only	67.7 (2)	23.3 (1.5)	9 (1.3)
Exp 4A	Not in recall	47.6 (1.9)	22.9 (0.8)	29.5 (1.7)
	Included in recall	64.1 (2.8)	23.2 (1.4)	12.8 (2.4)

Discussion

The cued recall procedure revealed different effects on category and exemplar memory of visual objects. Objects included in the cued recall phase were associated with better category memory than those not included in the cued recall. The enhanced category memory is consistent with previous studies that used maps (Carpenter & Pashler, 2007), colours (Sutterer & Awh, 2016), and objects (Maxcey & Woodman, 2014). The finding falls in line with the testing effect. This enhancement may also be attributed to the nature of the intervening recall. The cue used in that phase was categorical, which re-exposed participants to the category information of the objects. Thus, it is ambiguous whether the enhanced category memory is due exclusively to the testing effect or additionally to a re-exposure of the category label. This ambiguity does not affect the interpretation of exemplar memory because the cue provided no information about which exemplar was old.

The reduction in exemplar memory suggests that recounting one's memory can be detrimental, which outweighs the benefit of increased retrieval effort. The source of the detrimental effects may lie in verbal re-coding or the introduction of misinformation from inaccurate recall. The next experiment presents a replication of these findings along with an initial test on the role of verbal re-coding due to verbalization.

Experiment 2

Perhaps the most surprising finding of Experiment 1 was the lack of a positive exemplar memory benefit for objects recounted in the intervening phase. If anything, exemplar memory was impaired for the recounted objects. This impairment may have been due to verbal re-coding from verbally recounting the objects during the intervening recall phase. Experiment 2 aims to replicate this finding and examine the role that verbalization plays in producing this effect on exemplar memory. In Experiment 2, we investigate whether the lack of a positive effect, from recounting, on exemplar memory extends to objects that are recalled but not verbally described as well. To this end, we included two types of trials in the intervening phase. In one type, participants were asked to visualize the cued object and type out a verbal description, as before. In the second type, participants were only asked to visualize the cued object, minimizing verbal re-coding. As in Experiment 1, the final test was administered on the same day, and it involved a 4-alternative-forced-choice (4AFC) recognition task.

Method

Participants. Twenty-four participants, 16 females and 6 males, completed Experiment 2 (M = 19.1 years old, SD = 1.0).

Design and procedure. This experiment was the same as Experiment 1C, except for Phase 2. Specifically, of the 150 encoded objects, a random subset of 100 was included in Phase 2. On each trial of the cued recall task, participants were cued with the object name and asked to visualize the object they saw from the encoding phase. To increase the accountability of visualization, participants were asked to indicate whether they were highly confident, somewhat confident, or not confident that they had visualized the exact picture shown to them earlier. For a random half of these 100 objects, a window with the name of the object appeared after the confidence rating, and participants typed out a description of the object to enable someone else to identify it. For the other half of the objects, a verbal description was not requested. Instead, the next cue was presented immediately after the confidence rating. Thus, all 100 objects were visualized, and a random half of these were also described verbally. The other 50 objects were not included in Phase 2. Assignment of stimuli to conditions was randomized and counterbalanced across participants.

Results

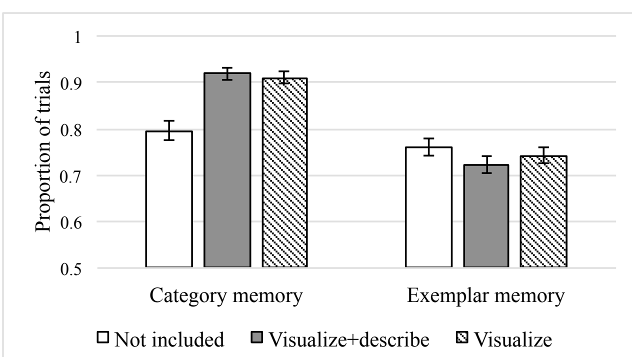

Figure S2.3. Recognition memory results from Experiment 2. Objects were separated based on whether they were visualized, visualized and described, or not included in Phase

2. Error bars show ±1S.E. of the mean.

Replicating Experiment 1, an ANOVA on recall condition (not included, visualized, or visualized+described) and memory type (category or exemplar) revealed a significant interaction, $F(2, 46) = 15.87$, $p < .001$, $\eta_p^2 = .41$. As shown in Figure S2.3., category memory was enhanced by intervening recall, $F(2, 46) = 30.60$, $p < .001$, $\eta_p^2 = .57$. It was about 12% better for both objects visualized and described, and those visualized only, than objects not in recall, $t(23)$s > 6.19, ps < .001. The former two conditions did not differ from each other, $t(23) = 0.57$, $p > .50$. In contrast, exemplar memory was slightly reduced by intervening recall, though this difference did not reach significance, $F(2, 46) = 1.59$, $p > .20$. Objects visualized and described had a 3.7% decline in exemplar memory than those not in recall, a marginally significant effect, $t(23) = 1.87$, $p < .08$. The deficit was 1.7% for those visualized only; this was not significant, $t(23) = 0.76$, $p > .45$.

Discussion

Intervening cued recall once again facilitated category memory but not exemplar memory. This pattern held both for objects visualized and described, and those visualized only. The similarity between these two conditions suggests that the act of intervening retrieval, rather than verbalization, is the primary driver behind the findings. The robust facilitation of category memory is consistent with previously observed testing effect, or with the possibility that memory is enhanced because participants were re-exposed to the category information. Unlike several recent studies on this topic, however, the enhancement did not extend to exemplar memory. If anything, exemplar memory showed a small decline of about 3%, an effect that reached significance in Experiment 1 but not in

Experiment 2. This finding provides initial evidence that intervening cued recall may introduce multiple effects: an increase in retrieval effort may be overridden by interference effects from the act of intervening retrieval. Due to the uniqueness of this finding in the recent literature on visual memory, we attempted a third replication in Experiment 3 using a simpler recognition task.

Experiment 3

The first two experiments used a 4AFC procedure to probe recognition memory. Exemplar memory was computed by restricting the analysis to trials in which participants chose the correct category. Although this design allowed us to simultaneously gauge category memory and exemplar memory, the presence of new categories during testing may have introduced extraneous effects. Experiment 3 aims to replicate the key findings of the first two experiments regarding exemplar memory. To this end, we simplified the final testing into a 2AFC procedure, in which only the old object and an exemplar foil were included.

Method

Participants. Twenty-four new participants completed Experiment 3. There were 13 females and 11 males (*M* = 19.8 years old, *SD* = 1.7).

Design and procedure. This experiment was the same as Experiment 1C except for a change in the final testing phase. Instead of being presented with four options, on each trial participants saw two options presented side by side on the horizontal meridian (the center of each image was 7.9º away from fixation). One option was an old object they saw earlier, the other was an object from the same category (i.e., exemplar foil). The participants' task was to click on the object they saw before.

Results and Discussion

Similar to the first two experiments, we did not observe a positive testing effect (Figure S2.4.). Exemplar memory for objects recounted was 2.8% lower than those not recounted, a difference that was not significant, $t(23) = 1.33$, $p > .19$.

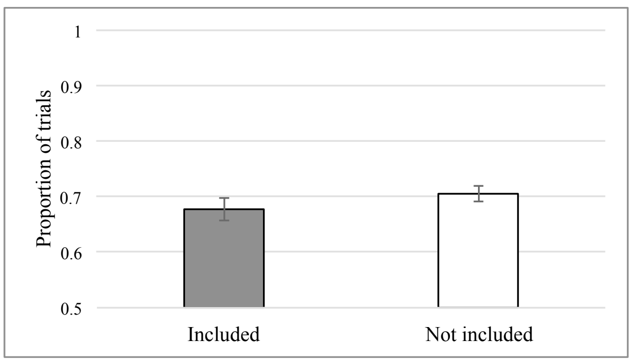

Figure S2.4. Recognition exemplar memory results from Experiment 3 for objects included and not included in the intervening recall phase. The final test included two choices: the old object and an exemplar foil. Error bars show ±1S.E. of the mean.

Together, the first three experiments showed an absence of the typical testing effect for exemplar memory. This was not simply a null result. The sample size was large - 85 combined across all experiments, a sample size adequate to detect a testing effect comparable in size to Sutterer and Awh (2016; 73 participants across three experiments). Exemplar memory was numerically lower for objects included in the intervening recall than for those not included. This reached significance in one experiment. This finding contrasts with the memory enhancement observed recently for maps (Carpenter & Pashler, 2007), colours (Sutterer & Awh, 2016), and objects (Maxcey & Woodman, 2014).

Experiment 4

The lack of a testing effect from the first three experiments may suggest that the retrieval effort account does not apply to exemplar memory. However, the finding is also compatible with the possibility that a testing effect was offset by interference from verbal re-coding or misinformation. One way to test this possibility is to increase the memory delay between Phases 2 and 3. The bifurcation model (Kornell, Bjork, & Garcia, 2011) suggests that when the final test is delayed, the testing benefit is larger because retrieval practice strengthens memory of the practiced items, protecting them from delay-induced forgetting. This is exactly the pattern seen in previous studies using verbal materials (Liu & Reder, 2016; Roediger & Karpicke, 2006; Wheeler, Ewers, & Buonanno, 2003). Here we investigated whether the same is true with exemplar memory.

Two experiments were administered. In both experiments the delay between Phases 2 and 3 was increased to 24 hours. Experiment 4A was identical to Experiment 1C in its use of a 4AFC final test; Experiment 4B was identical to Experiment 3 in its use of a 2AFC final test. These experiments were conceptual replications of each other. If a testing effect on exemplar memory is observed with the delay, it would suggest that intervening recall could have long-term faciliatory effects on memory.

Method

<u>Participants.</u> The 24 participants in Experiment 4A included 19 females and 5 males (*M* = 19.6 years old, *SD* = 1.7). The 24 participants in Experiment 4B included 16 females and 8 males (*M* = 20.3 years old, *SD* = 2.1).

<u>Design and procedure.</u> Other than a change of the delay interval between Phases 2 and 3 to 24 hours, Experiment 4A was identical to Experiment 1C, whereas Experiment 4B was identical to Experiment 3. The final memory test was a 4AFC in Experiment 4A, permitting separate assessment of category and exemplar memory. The final memory test was a 2AFC in Experiment 4B, a simpler procedure focusing exclusively on exemplar memory.

Results

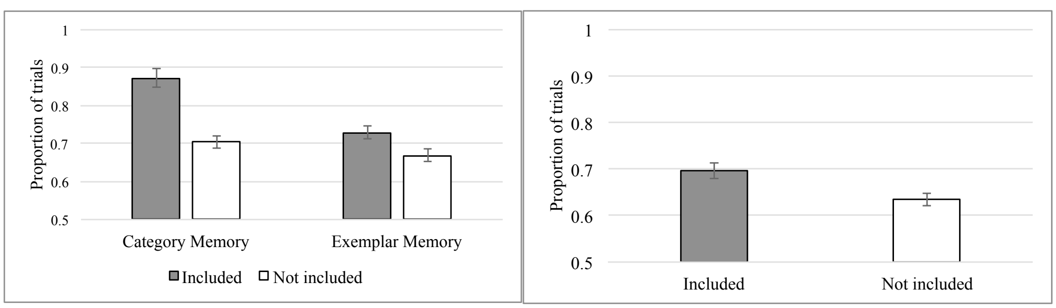

Figure S2.5. Recognition category and exemplar memory results from Experiment 4A (left-4AFC task) and recognition exemplar memory results from 4B (right-2AFC task) for objects included and not included in the intervening recall phase. The memory delay between Phases 2 and 3 was 24 hours. Error bars showed ±1S.E. of the mean.

Following a 24-hour delay between the intervening cued recall and the final test, we observed a strengthening of the testing effect in both Experiments 4A and 4B (Figure

S2.5.). To enable a direct assessment of the modulating effect of delay, the following analysis included delay interval as a between-subject factor.

1. <u>4AFC: immediate (Experiment 1) vs. 24-hr delay (Experiment 4A)</u>

Category memory was better for objects included in the intervening recall than for objects not included. This was true both when the final test was immediate and when it was delayed by 24 hours, though the effect was stronger in the latter case. An ANOVA with intervening recall as a within-subject factor and memory delay as a between-subject factor showed a significant main effect of intervening recall, $F(1, 59) = 251.18$, $p < .001$, $\eta_p^2 = .81$, as well as a significant interaction, $F(1, 59) = 5.66$, $p < .05$, $\eta_p^2 = .09$. Intervening recall enhanced category memory by about 16.8% with a 24-hour delay ($t(23) = 12.89$, $p < .001$), compared to 12% without a delay.

Adding a 24-hr delay flipped effects of intervening recall on exemplar memory. Whereas objects included in the intervening recall yielded 3% worse exemplar memory than objects not included (Experiment 1), this difference switched signs to a 6% gain following a 24-hour delay (Experiment 4A; $t(23) = 3.87$, $p < .01$). An ANOVA on intervening recall and memory delay revealed a significant interaction, $F(1, 59) = 18.58$, $p < .001$, $\eta_p^2 = .24$. Importantly, the change from negative to positive effects of recounting was mainly due to a decrease in exemplar memory for items not included in recall from 75.6% (S.E. = 1.6%) when there was no delay to 66.9% (S.E. = 1.6%) when there was a 24-hour delay after intervening recall, $t(59) = 3.73$, $p < .001$. Memory for items included in recall was similar when there was no delay (M = 72.3%, *S.E.* = 1.8%) and when there was a delay (M = 72.9%, *S.E.* = 1.7%), $t(59) = 0.22$, $p = .83$.

2. 2AFC: immediate (Experiment 3) vs. 24-hr delay (Experiment 4B)

Replicating the findings from the 4AFC, in the 2AFC task we again found opposite effects of intervening recall for immediate and delayed memory tests. The 2.8% numerical cost of exemplar memory for objects included in the intervening recall (Experiment 3) reversed to a 6.2% facilitation after a 24-hour delay (Experiment 4B, $t(23) = 3.80$, $p < .01$). An ANOVA on intervening recall and memory delay yielded a significant interaction, $F(1, 46) = 11.38$, $p < .01$, $\eta_p^2 = .20$. This result was also due mainly to a decline in exemplar memory for items not included in recall from when there was no delay ($M = 70.5\%$, $S.E. = 1.5\%$) to when there was a delay ($M = 63.5\%$, $S.E. = 1.4\%$), $t(46) = 3.43$, $p < .01$. There was no significant difference for items included in recall when there was no delay ($M = 67.8\%$, $S.E. = 2.1\%$) compared to when there was a 24-hour delay after intervening recall ($M = 69.7\%$, $S.E. = 1.7\%$), $t(46) = 0.72$, $p = .48$.

Discussion

Inserting a 24-hour delay between intervening recall (Phase 2) and final recognition (Phase 3) produced a more positive effect of intervening recall. The already positive effect on category memory increased in strength following a 24-hour memory delay. The small negative effect of intervening recall on exemplar memory reversed to a positive testing effect. These findings are consistent with previous research that used verbal stimuli. Time delay appears to enhance the benefit of intervening memory retrieval (e.g., Liu & Reder, 2016). Importantly, adding a time delay uncovers the positive testing effect that was otherwise not detectable in Experiments 1-3. These findings suggest that intervening recall likely slows forgetting of recalled items such that a clear testing benefit on exemplar memory emerges only when a delay follows intervening recall test.

General Discussion

Four experiments examined effects of intervening retrieval on visual memory. Using a recognition test that isolates exemplar memory from category memory, we showed that intervening retrieval facilitates exemplar memory when tested 24-hours later. The enhancement is reminiscent of the testing effect typically seen with verbal materials. However, when testing is administered immediately, no facilitation is seen. If anything, exemplar memory for recounted objects is slightly reduced. This latter finding suggests that intervening retrieval may introduce detrimental effects on memory. The detrimental effects are anticipated from prior work on the verbal overshadowing effect (Schooler & Engstler-Schooler, 1990) and misinformation effect (Loftus, 2005). Thus, exactly how intervening retrieval affects exemplar memory depends on the relative strength of opposing effects. Temporal delay modulates the competition between different effects. At longer delays, the benefit of increased retrieval effort increases (e.g., Roediger & Karpicke, 2006), producing a net gain. At shorter delays, the detrimental effects of verbal re-coding and misinformation may sometimes outweigh the positive testing effect.

The temporal dynamics revealed in this study are consistent with previous findings using word lists (Liu & Reder, 2016) or passages (Roediger & Karpicke, 2006). It supports the bifurcation model, according to which an increase in retrieval effort slows down forgetting (Kornell et al., 2011). The lack of a testing effect when tested immediately, however, contradicts recent studies on visual memory, most of which tested memory immediately after retrieval practice (e.g., Carpenter & Pashler, 2007). Several factors may contribute to this contradiction. First, previous studies did not fully distinguish category memory from exemplar memory. The old-new memory test, for example, can be supported by both. Our study suggests that category memory may be less vulnerable to detrimental effects of verbal re-coding and misinformation. Second, the studies differ in the nature of the intervening retrieval task, which may have influenced both the benefit from retrieval effort and the cost from verbal re-coding and misinformation. Intervening tasks that yield highly accurate retrieval, and minimize verbal re-coding and the introduction of misinformation, are likely to produce just a positive testing effect. Carpenter and Pashler (2007) used an altered map to probe memory for missing features. The similarity of the cue to the actual map is likely to minimize inaccurate retrieval. Maxcey and Woodman (2014) used a 2AFC recognition task as the retrieval practice, which was also unlikely to introduce false information. In contrast, the open-ended visualization and description task used in the current study is likely to yield inaccurate retrieval. In fact, an examination of participants' verbal recall in Phase 2 suggests that their recall included misinformation. Coders, as a result, were more likely to choose the wrong exemplar when participants themselves chose the exemplar foil, suggesting the presence of false memory. This finding is consistent with previous work on false memory, which shows that exposure to items either verbally (Neushatz, Benoit, & Payne, 2003) or visually (Sapkota, van der Linde, & Pardhan, 2015) can lead to the erroneous memory that a related, but not previously presented, item was also seen before during the experiment. In our study, details of a tent seen in the participant's past experiences (outside of the experiment) may have been included in their verbal recall of the tent shown during the first phase of the experiment. The inaccurate descriptions in the recall may, thus, have increased detrimental effects of the retrieval practice, overriding the positive testing effect. Having said this, the source of the false memory in this study is unclear - it could be that people had false memory to begin with, or that the intervening recall had introduced false memory. Future studies are needed to further examine the nature of object memory.

Similar to other studies on the testing effect, in our experiments participants could be multitasking during the intervening recall phase - recalling the cued objects as well as remembering all of the objects seen in phase 1. It is well known that concurrent tasks interfere with memory (Sapkota, Pardhan, & van der Linde, 2013). It is therefore possible that the act of recalling the cued objects had interfered with memory. However, any such interference effect should apply to all objects held in memory, not just to those recalled in Phase 2. Thus, generic dual-task interference does not explain why in some cases (e.g., Experiment 1) exemplar memory was impaired for only those objects that were recalled and not for other objects that were not recalled in phase 2.

Our study reconciles apparent contradictions from previous studies on visual memory. We suggest that retrieval effort, verbal overshadowing, and misinformation jointly affect how intervening retrieval affects visual memory. Temporal delay and the nature of retrieval practice affect the relative strength of these factors. The dynamic view of intervening retrieval explains a wide range of empirical findings, from detrimental effects of verbal overshadowing to the standard testing effect. Because memory for visual details is particularly vulnerable to interference from verbal re-coding and misinformation, recounting one's visual memory may yield interference effects not typically seen for verbal stimuli.

Our study has implications for practical situations involving visual memory, such as eyewitness memory, classroom learning of graph or images, and remembering visual experience from the past. Recounting one's visual memory can be beneficial, especially if the memory needs to be retained for a long time. However, such recounting may also introduce interference, which at shorter intervals may override the benefit of the retrieval practice. An important future research question is to identify approaches that minimize interference from verbal re-coding and misinformation.

The current study used category names to cue object memory during the intervening recall phase. As noted earlier, this procedure re-exposed participants to the category label, a factor that may have contributed to enhanced category memory for objects included in the recall. Memory of the cue could then have faded over the 24-hour delay, and contributed to changes in category memory over time. The use of a category cue, therefore, made it more difficult to interpret how intervening recall influenced category memory. This was the reason we focused on exemplar memory in the present study. Because the categorical cue was uninformative of which exemplar participants saw earlier within a category, effects of intervening recall on exemplar memory cannot be explained by the nature of the cue. Nonetheless, future studies should use non-categorical cues to investigate effects of intervening recall on visual memory.

Conclusion

Four experiments suggest that intervening retrieval affects visual memory in several ways. One is through facilitation from increased retrieval effort, and the other is through interference effects from a likely combination of verbal re-coding and misinformation effects. A longer delay strengthens the positive effects from intervening retrieval. These findings support the view that visual memory is dynamic. Despite its high capacity and fidelity, visual memory is not a snapshot of one's previous experience. Rather, it is constantly in flux and modified as we think or talk about what we have seen.

3. Section 3

The experiments in Section 2 revealed that intervening recall can enhance visual detail memory in the long run by slowing forgetting of the material. They tapped into "item" memory with the testing of individual objects presented in isolation. In the real world, however, items typically appear in a scene context (e.g., a bench in a park). Does the testing effect similarly affect both item and source memory? Section 3 examines this question by superimposing objects against a background scene. We investigated the effects of intervening recall on both item and source memory (i.e., which scene an object was from).

Section 3 also addressed one specific aspect of Section 2's design – the nature of the memory cue used for intervening recall. Instead of presenting a category cue which re-exposed participants to an object's category, in Section 3 we used the scene context as the cue for intervening recall and examined whether the testing benefit remained.

Finally, Section 3 provided an in-depth analysis on the nature of intervening recall to examine how much the testing benefit relates to "fruitful" retrieval during the intervening test. To this end, we classified participants' verbal descriptions into failed-to-recall, recalled-without-visual-details, and recalled-with-visual-details. Instead of treating all objects included in the recall phase as a single condition, we examined whether the nature of recall was related to how much memory enhancement occurred later on. Together, these experiments generalized and deepened the research from Section 2.

Data reported in this section are from an article in preparation for submission. The paper is reproduced in its entirety.

The generalizability of the testing effect across memory cues and attributes

Deborah H. Tan

Yuhong V. Jiang

Department of Psychology, University of Minnesota

Abstract

The benefits of intervening retrieval on subsequent verbal memory (i.e., the testing effect), have been widely studied. While we previously found that the effect can extend to visual material, the benefit was observed under specific conditions where the memory cue exposed participants to aspects of the encoded material. Here we examine the generality of the testing effect and ask whether 1) the testing effect extends to visual stimuli with a more neutral memory cue, and 2) whether the testing effect can be observed across different attributes (both item and source information) of a memory. We presented photographs of objects to participants, which were superimposed onto semantically-unrelated scenes. We then introduced an intervening recall phase with the scene as a memory cue. Both the item and its scene pairing had to be retrieved during recall. In the final memory test 24 hours later, we tested participants' memory for the objects (item memory), or the scenes (source memory) when presented an old object. As found previously, intervening recall enhanced subsequent item detail memory. Source memory was also facilitated by intervening recall. Fine-grained analysis of the amount of recall from participants suggests that the benefits retrieval has on subsequent memory are

related to how much elaboration occurs during intervening retrieval – lending support to the retrieval effort account of the testing effect.

Keywords: visual memory, intervening recall, testing effect, item memory, source memory

Introduction

The testing effect refers to the finding that attempts to retrieve a memory slow down its forgetting. In a classic demonstration, Karpicke and Roediger (2008) first trained participants to learn 40 Swahili-English word pairs. Participants then alternated between restudying the word pairs and being tested to recall English words when presented the Swahili words. Some participants repeatedly restudied the word pairs and were only tested on word pairs that they had unsuccessfully recalled before, while others were repeatedly tested on the word pairs and only restudied the word pairs they were not able to recall previously. In a final test one week later, Karpicke and Roediger found that participants who had repeatedly retrieved the items (i.e., during test practice) had better long-term retention than participants who repeatedly restudied the items instead. Hence, repeated testing slows forgetting of the items and enhances long-term retention. The beneficial effect of intervening memory retrieval, or "testing", has also been observed with word lists (Carpenter & DeLosh, 2006), passages (Roediger & Karpicke, 2006), general knowledge (Carpenter, Pashier, Wixted, & Vul, 2008), and face-name pairs (Carpenter & DeLosh, 2005). The leading theory of the testing effect is the retrieval effort account, according to which retrieval effort during intervening testing increases the accessibility of the encoded material, thereby improving subsequent memory for that material (Carpenter & DeLosh, 2006; Carpenter, 2009). By this logic, the testing effect should appear in the long term as long as the retrieval effort exerted during intervening testing is sufficient. This account suggests that the testing effect should be generalizable across many types of stimuli, such as visual stimuli, and when various aspects of the original memory are tested.

Several recent studies provide evidence that the testing effect applies to visual stimuli. Even though visual memory is not readily describable, and talking about one's visual memory (such as a face one witnessed earlier) can interfere with visual memory (Schooler & Engstler-Schooler, 1990), the testing effect manifests in visual stimuli. In Carpenter and Pashler (2007), participants first studied two maps. Next, one map was studied for a second time (the "restudy" condition), while the other required participants to recall missing features (the "testing" condition). Finally, memory for both maps were tested. The testing condition yielded better memory of the map than the restudy condition. A similar finding was observed in Sutterer and Awh (2016), in which participants were exposed to animal shapes in various colors. Next, participants either re-studied these stimuli again, or engaged in memory retrieval by choosing each shape's color on a color wheel. In a later memory test, participants recalled more colors for the tested shapes than for the restudied shapes.

The testing effect also extends to photographs of common objects. In Tan and Jiang (2018; see Section 2), participants were first presented 150 objects for 1s each. Next, they were presented with the basic-level category names of a subset of the encoded objects and asked to type out a verbal description for those objects seen in the preceding phase. Either 2 minutes or 24 hours after intervening recall, participants were tested on their recognition memory for all 150 objects in a final forced-choice task. When the delay was 2 minutes, intervening recall enhanced category memory, but impaired exemplar (i.e., detail) memory. As the delay increased to 24 hours, intervening recall enhanced both exemplar and category memory.

Although intervening retrieval strengthens visual memory, especially when tested after a long delay, the effect has so far been revealed under specific conditions. In Tan and Jiang (2018), intervening retrieval was cued based on the category of the objects. This method re-exposes participants to the objects' category, confounding the testing effect with restudying. This was a concern when category memory was assessed – better category memory for objects included in the recall phase can be attributed either to the testing effect, or to re-exposure to the object category. This raises questions about how intervening retrieval affects visual memory when a more neutral type of cue is used, which eliminates re-exposure.

Another restriction of previous research on visual stimuli is their focus on one type of memory – item memory. Neglected in this research is source memory – the contextual information that accompanies a specific visual object. In everyday life, objects rarely occur in isolation; they tend to appear in the context of other objects or background scene or at a specific time or event. The presence of both item and source memory raises the question of whether the testing effect applies to multiple aspects of a memory, and if so, whether the effect is qualitatively similar for item memory and source memory. In other words, is the testing effect attribute-independent?

Several studies suggest that item and source memory can be partially dissociated. Spencer and Raz (1994) showed that aging has a more deleterious effect on source memory than on item memory. They asked healthy young (18-35 years old) and older (65-80 years old) participants to remember statements about well-known or fictitious people. The statements were presented on pink or blue cards and delivered in room A or room B. Subsequently, the researchers tested participants' memory of the exact statement (item memory) and the color or room associated with the statement (source memory). Compared with young adults, older adults' memory for the statements was similar, but their memory for the source of the statement was significantly worse.

Other studies (Glisky, Polster, & Routhieaux, 1995; Davachi, Mitchell, & Wagner, 2003; Fan, Snodgrass, & Bilder, 2003) provided neuropsychological evidence that item and source memories rely on different parts of the brain, and thus may have different underlying mechanisms. In Glisky et al. (1995), older adults were asked to remember sentences spoken in either a male or female voice. Later their memory for both the sentences (item memory) and the associated speaker gender (source memory) were assessed. Glisky et al. divided participants in two different ways: high vs. low medial temporal lobe function based on their performance on four memory tests typically associated with medial temporal regions of the brain, or high vs. low frontal function based on their performance on five tests typically used to measure frontal lobe function (e.g., mental arithmetic from the WAIS-R). Results showed that item memory was better in older adults with high rather than low medial temporal lobe function, but source memory did not differ between the two groups. In contrast, source memory was better in older adults with high rather than low frontal lobe function, but item memory was comparable between the two groups. This study suggests that the medial temporal lobe and frontal lobe contribute differently to item and source memory, raising the question of whether intervening testing will differentially affect each type of memory.

While differential effects of intervening retrieval on item and source memory are a possibility based on Glisky et al. (1995)'s finding, a study by Staresina and Davachi (2009) suggests that source memory may not rely solely on the frontal lobe. The researchers examined the role of the hippocampus in the binding of elements in memory and found that the hippocampus is involved in associative binding, particularly when the different elements (e.g., the object and its color) are separated spatially from each other. Their results suggest that, contrary to Glisky et al. (1995)'s finding, the medial temporal lobe is important for binding a source to its corresponding object in memory, and thus, would be important for source memory as well.

Other neuroimaging studies have examined brain mechanisms underlying the testing effect specifically, including a test of the retrieval effort account (Carpenter & DeLosh, 2006). These studies have found that successful retrieval of material during intervening tests strengthens the neural representations of that material in the temporo-parietal regions (see van den Broek et al., 2016 for review). Other studies have found an involvement of the frontal lobe, particularly the Ventral Lateral Prefrontal Cortex, during intervening testing compared to restudy (e.g., van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013; Wing, Marsh, & Cabeza, 2013). The frontal lobe may be involved to suppress inaccurate competing information that is retrieved (see van den Broek et al., 2016). Thus, both the frontal and temporal regions are involved in the testing effect. On that note, without directly investigating the effects of intervening testing on item and source memory, no clear conclusions about whether intervening tests have similar effects on item and source memory can be drawn from the past neuropsychological and neuroimaging studies.

Using word lists as stimuli, one recent study found evidence of a testing effect on source memory. In Brewer, Marsh, Meeks, Clark-Foos, and Hicks (2010), participants listened to two word lists, where words in each list were spoken by either a male or a female speaker. Immediately afterwards, half of the participants were asked to recall the words and the gender of the speaker for each word. The other half completed an unrelated filler task. In a final recognition test, participants were shown one word at a time and asked whether it was previously spoken in a male or a female voice. Brewer et al. found enhanced memory for participants who previously recalled the words and the speaker's gender, compared to those who did an unrelated task. This testing effect depended on the recall of speaker gender in the intervening phase. In a control experiment, participants were only asked to recall the words, but not the gender of the speaker, during the intervening phase. In this experiment, no memory enhancement for the speaker's gender was found. Thus, much like item memory, source memory appears to benefit from effort being exerted to retrieve that specific type of memory in an intervening phase. Brewer et al., however, did not examine how intervening recall affected item memory. Hence, it is still unclear whether the testing effect is indeed attribute-independent.

The main goal of the present study is to examine the generality of the testing effect on visual memory. We address two questions. First, is visual memory enhanced by intervening testing, when there is no re-exposure of the category information during that phase? Second, how does intervening testing affect item memory and source memory for visual stimuli? To this end, we ran two experiments. Experiment 1 examined the effects of intervening recall on an object's exemplar and category memory (i.e., item memory). Experiment 2 examined the effects of intervening recall on source memory.

In these experiments, participants were first shown famous scenes – one at a time – each with distinct objects superimposed onto them. Scenes were semantically unrelated to the objects. In a second intervening recall phase, participants were presented half of the scenes and were asked to type out a description of all the objects that were presented on those scenes in the previous phase. As such, the scenes served as memory cues for the intervening recall task. This procedure did not re-expose participants to category information during the intervening phase, a crucial difference from Tan and Jiang (2018). In order to retrieve the objects when presented a scene, participants would have to retrieve both the scene and its objects to complete recall. This task asks that participants retrieve both item and source memory during intervening recall, making retrieval demands for item and source memory similar to each other.

The final recognition test examined either item (Experiment 1) or source memory (Experiment 2). In Experiment 1, four objects were displayed for a four-alternative-forced-choice task (4AFC): the old object, a different exemplar from the same basic-level category as the old object (i.e., within-category foil), and two exemplars from a new basic-level category not seen before (i.e., between-category foils). Participants had to choose the old object. With this experimental design, we aimed to examine whether intervening recall enhances object memory (Tan & Jiang, 2018) even when the memory cues did not provide any information about the object categories.

The final recognition test of Experiment 2 assessed whether the effects of intervening recall on source memory are qualitatively similar to that for item memory (i.e., intervening testing enhances subsequent memory). To this end, we presented each old object one at a time and asked participants to select the scene on which the object was shown in the first encoding phase.

Past studies have shown that the benefits from intervening recall depend, to some extent, on accuracy of recall during that intervening phase (Liu & Reder, 2016; Kornell, Bjork, & Garcia, 2011; Rowland, 2014). More accurate recall tends to be associated with larger benefits of intervening recall on subsequent memory performance. While useful to examine how the accuracy of verbal descriptions produced during recall are related to subsequent memory performance, classifying verbal descriptions as being accurate or inaccurate was challenging. Descriptions were rarely completely accurate or completely inaccurate. There was often a mix of accurate and inaccurate information, making coding by accuracy difficult and subjective (see Tan & Jiang, submitted; Section 4). Another way of analyzing verbal recall was to focus on the amount of retrieval that occurred during the intervening phase. According to the retrieval effort account, the more material that is retrieved during a test, assuming retrieval is successful, the more elaboration of the material there is, which increases the number of retrieval routes to the material (van den Broek et al., 2016; Carpenter, 2009) and enhances subsequent memory. This definition has been supported by several studies showing that intervening *recall* tests produce greater long-term retention of material than intervening recognition tests (Bjork & Whitten, 1974; Kang et al., 2007). On that note, in both our experiments, we classified the verbal descriptions from participants according to the amount of "fruitful" retrieval that occurred. For objects that were included in recall, we distinguish among those that participants failed to recall (no retrieval), those where only the basic-level category was recalled without any elaboration (small amount of retrieval), and those that were recalled with additional details (e.g., color of the object - greater amount of retrieval). This finer-grained analysis of the status of retrieval allows us to more closely examine the retrieval effort account, and thus, the generality of the testing effect.

Experiment 1

This experiment examines whether intervening recall facilitates later visual item memory and serves as a conceptual replication for Tan and Jiang (2018). Unlike Tan and Jiang (2018), the intervening recall was prompted with a semantically-unrelated contextual cue that does not re-expose participants to aspects of the visual objects. To this end, participants encoded objects in the context of background scenes. In Phase 1 (encoding), nine categorically-distinct objects were overlaid on a famous scene. There was a total of 72 objects, randomly assigned to 8 scenes. Because the objects were randomly assigned, they did not semantically cohere, neither did they relate to the background scene. Participants were asked to remember the objects and the scene on which they were presented.

In Phase 2 (intervening cued recall), 4 of the scenes were presented, one at a time, to serve as retrieval cues. Upon seeing the scene, participants had to type out a description of all the objects that had previously been presented on the scene shown. The arbitrary pairing of the scenes and objects meant that any testing effect derived from Phase 2 would be attributed to the retrieval practice, rather than to re-exposure to aspects of the objects.

In Phase 3 (final recognition memory test), item recognition memory was assessed for all the objects that were shown in Phase 1. Each trial displayed four images against a blank background: the old object, a within-category foil, and two between-category foils. In line with past findings (Liu & Reder, 2016; Roediger & Karpicke, 2006; Wheeler, Ewers, & Buonanno, 2003), Tan & Jiang (2018) found a greater testing effect when there was a 24-hour delay between intervening recall and the final test.

Therefore, in this experiment, while Phases 1 and 2 were administered on the same day, Phase 3 was administered 24 hours after Phase 2. We examined whether memory for objects included in Phase 2 was better than memory for objects not included in Phase 2.

The use of both category and exemplar foils enables us to compute category and exemplar memory. Category memory was calculated as the proportion of times that participants chose the correct category (the old object or the within-category foil). Exemplar memory was calculated as the proportion of times that participants chose the old object, given that the correct category was chosen (i.e., P(old object | correct category)).

A further design feature of Experiment 1 is the analysis of whether there was a relationship between what occurred during intervening recounting and the participants' subsequent recognition memory performance. In Experiment 1 two raters separated objects cued in Phase 2 into three groups: objects that participants failed to recall, objects that participants recalled with just a category name and no elaboration, and objects that participants described with visual details. The retrieval effort account predicts that the objects recalled with visual details should also show the greatest memory enhancement in the subsequent recognition memory test.

Method

Participants. Participants in all the experiments were students from the University of Minnesota between 18-35 years of age. They were native English speakers, naive to the purpose of the study, had normal or corrected-to-normal visual acuity, and normal color vision. Participants signed an informed consent and were compensated with extra course credit.

Sample size. Sample size was determined based on a number of factors. A sample size of 24 in Experiment 4 of Tan and Jiang (2018), which also had a 24-hour delay between intervening recall and final recognition testing, was adequate to detect a testing benefit from intervening recall for both category and exemplar item memory. In addition, to counterbalance the experimental materials across the conditions in this study, the sample size needed to be in multiples of 16. Given these considerations, we aimed to test 32 participants in Experiment 1and 32 participants in Experiment 2. Final sample size fluctuated slightly due to participant signups.

Experiment 1 had 34 participants, 26 females and 8 males (*M* = 19.9 years, *SD* = 2.3 years).

Materials. Participants were tested individually in a room with normal interior lighting. The program was written in MATLAB (www.mathworks.com) and Psychtoolbox (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a 19” CRT monitor (1024x768 pixels).

Eight photographs of famous scenes around the world were chosen from Google Images. They were scenes of the following places: Statue of Liberty, Great Wall of China, Mount Rushmore, Eiffel Tower, Taj Mahal, Cloud Gate, Stonehenge, and Sydney Opera House. We used famous scenes to minimize the need to learn new scenes, and to increase the distinctiveness of the scenes and their effectiveness as retrieval cues.

Photographs of common objects were sampled from Brady and Oliva’s object database (http://cvcl.mit.edu/MM/stimuli.html). The stimulus set contained 144 distinct basic-level categories, each with two exemplars. 72 categories were used during the encoding phase. Nine objects were randomly assigned to each scene, reducing the semantic association between the scene and the objects. This design is comparable to previous verbal studies. In Spencer and Raz, (1994), for example, pink and blue cards were paired with semantically-unrelated words. The arbitrary pairing of objects and scenes also prevented participants from using schema-aided memory (Hess & Slaughter, 1990). This minimizes the re-exposure to aspects of the visual objects when the scenes are used as retrieval cues. In order to recall the correct object, participants would have to retrieve the actual pairing of the scene with that object, instead of making guesses based on semantics.

Procedure and Design. Participants were tested in three phases, separated by two minutes between Phases 1 and 2, and 24 hours between Phases 2 and 3 (Figure S3.1.).

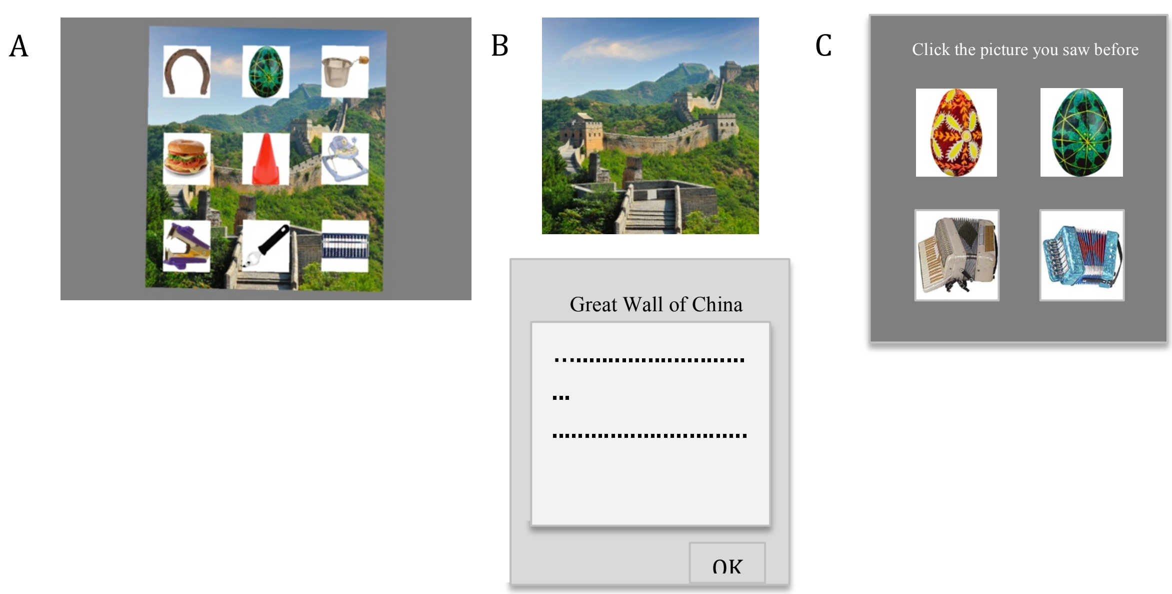

Figure S3.1. A schematic illustration of the encoding, cued recall, and final memory test phases. A. The encoding phase presented eight different scenes with nine categorically *distinct objects on top of each scene at a pace of 2 mins/presentation. B. An example of the intervening cued recall phase. Participants were shown half of the scenes, one at a time, and asked to type out a description of all the objects that were presented on that scene in the preceding phase. C. The recognition phase with an old object, a within-category foil, and two between-category foils not seen previously. 24 hours separated the intervening cued recall phase and the final recognition test.*

(1) Familiarity phase. In order to ensure that all participants were familiar with the eight famous scenes before the encoding task, participants were presented the eight scenes, one at a time, and asked to name them. They were shown all 8 scenes repeatedly until they were able to name all eight scenes. Once they were able to do so, they proceeded to the main experiment.

(2) Phase 1: Encoding. In the first phase, participants viewed eight different scenes (scene size: 34.8ºx34.8º) with nine categorically distinct objects (image size: 7.7ºx7.7º) superimposed on top of each scene. Each scene and its objects were displayed for two minutes each. Participants were asked to remember what the objects looked like and on which scene they were placed. They were also told that at a later time, they would be asked to recall the objects from the scenes. Across the 8 scenes, there were a total of 72 objects presented in Phase 1. These objects were randomly selected from the 144 categories, with the remaining 72 categories of objects used as foils in the final memory test. Which set was used for encoding and which as foils were counterbalanced across participants. In addition, to control for stimulus differences, we randomly divided the objects into groups of 9 objects. The scene on which each group of 9 objects were presented was counterbalanced across participants. In addition, for a given object

category, the exemplar that was encoded in Phase 1 was counterbalanced across participants. The other exemplar which was not encoded served as a within-category foil in the final memory test phase.

(3) Phase 2: Intervening cued recall. Half of the scenes encoded (N=4) were used in the second phase. In this phase, participants were shown a famous scene presented previously, and were asked to recall the objects that were presented on that scene in the previous phase. They were asked to type out as many details as possible about the objects that were on that scene. After they finished typing, they clicked “OK” to proceed to the next trial. No feedback was given. The assignment of the stimuli for inclusion in Phase 2 was counterbalanced across participants, meaning that a given scene, such as the Sydney Opera House, was presented as a memory cue for half of the participants, and not presented for the other half. The extensive randomization and counterbalancing used in this experiment ensured that, across participants, a given object is equally likely to be an encoding object or a memory foil, and equally likely to appear against any of the 8 background scenes. In addition, a given encoded object is equally likely to be included or excluded from Phase 2’s recall. The 36 objects paired with the 4 scenes are considered “in recall”, whereas the 36 objects paired with the 4 scenes not displayed in Phase 2 are considered “not in recall.”

(4) Phase 3: Final memory test. The final memory test occurred 24 hours after Phase 2. It included 72 trials. On each trial, four objects were presented, one in each quadrant (image size: 14.7°x14.7°; the center of each image was 11.7° from fixation, horizontally displaced by 7.9° and vertically displaced by 8.6° from fixation). One object, the old object, was previously presented in the encoding phase (e.g., a suitcase). Another

object was from the same basic-level category as the old object, but was a new exemplar (within-category foil; e.g., a different suitcase). The other two objects were two exemplars from a basic-level category not presented previously (between-category foils; e.g., two toothbrushes, and toothbrush was never presented in the experiment). The positions of the four objects on the screen were random. Participants were asked to click on the object that they had seen before. A tone provided feedback about their accuracy.

Results

On every trial of the final memory test, the participants were shown an old object, a within-category foil, and two between-category foils in a 4AFC task. Participants' responses could fall into any of these three options. This allowed us to compute category memory and exemplar memory separately (see Method).

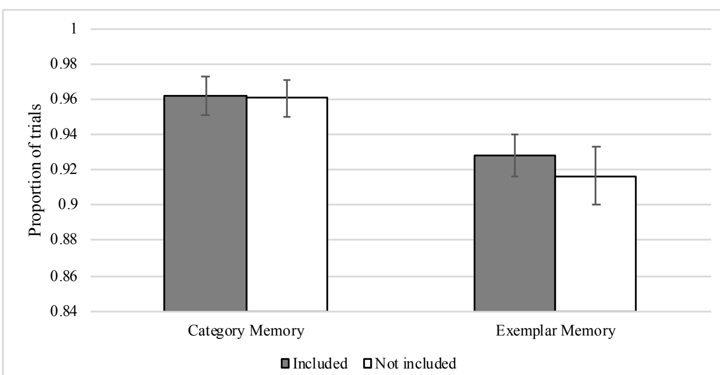

Figure S3.2. Recognition item memory results from Experiment 1 for objects included and not included in the intervening recall phase. Error bars show ±1S.E. of the mean.

In the first analysis, we combined all 36 objects in recall as one condition, and all 36 objects not in recall as another condition (Figure S3.2.). An ANOVA on type of memory (category or exemplar) and intervening recall (included or not included) revealed only a significant main effect of memory type, $F(1, 33) = 25.93$, $p < .001$, $\eta_p^2 = .44$. Participants were more accurate in identifying the category of the object than in identifying the exact exemplar. However, there was no main effect of intervening recall, $F(1, 33) = 0.81$, $p = .37$ and no interaction, $F(1, 33) = 0.71$, $p = .41$. Objects included in recall were not remembered better than those not in recall, for either category or exemplar memory.

The above analysis treated all 36 objects paired with the 4 retrieval cues (scenes) equivalently, regardless of retrieval success. This analysis blurred any effects of retrieval success across the different objects. Because not all objects were associated with fruitful retrieval in Phase 2, the mere presence of a cue that could have led to their retrieval may not have influenced memory.

To further examine whether what occurs during intervening testing is associated with subsequent memory performance, we separated the 36 objects included in retrieval into three types. Specifically, two raters determined whether objects previously paired with each scene cue were 1) RecalledVisual, i.e., recalled with visual details (e.g., “yellow watch”), 2) RecalledNoVisual, i.e., recalled without visual details (e.g., “watch”), or 3) FailedToRecall, i.e., not recalled when it should have been (e.g., a watch was paired with the Eiffel Tower, but the participant did not recall there was a watch when shown the Eiffel Tower). These categories were contrasted with 4) NotInRecall, i.e., objects not cued for recall, e.g., an object was paired with a scene that was not shown during the intervening recall phase. Some objects were left out from this classification. Specifically, 2.5% of the objects that were intrusions, i.e., an object that was recalled but was previously paired with a scene that had not been shown during the intervening recall phase. Because they occurred infrequently, they will not be discussed further. Inter-rater reliability was high (0.97 in both Experiments 1 and 2). Hence, we used the data from only one of the raters in subsequent analyses. Table S3.1. shows the average number of trials for each type of recall in Experiments 1 and 2.

Table S3.1. Mean proportion of each type of recall (FailedToRecall, RecalledNoVisual, RecalledVisual) out of the 36 objects that were cued in Phase 2 (intervening recall) in Experiments 1 and 2. Standard error of the mean is shown in parenthesis. The three types of recall add up to a total proportion that is slightly less than 100% because some of the cued objects were classified as intrusions.

	FailedToRecall	RecalledNoVisual	RecalledVisual
Exp 1	61.4 (4.5)	16.3 (2.9)	18.7 (3.7)
Exp 2	67.7 (4.0)	13.1 (3.0)	15.5 (2.8)

It should be noted that the raters did not specifically code for the accuracy of the recalled details. This is both because what participants reported could be ambiguous, and also because we were mainly interested in the amount of visual details as a measure of fruitful retrieval. In addition, the small number of stimuli prevented us from further dividing the "recall with visual" into smaller categories (e.g., recall with completely accurate visual details, recall with some accurate details, and recall with inaccurate visual details). How accuracy of the intervening recall descriptions affects subsequent memory is analyzed in a separate study described in Section 4 of the dissertation.

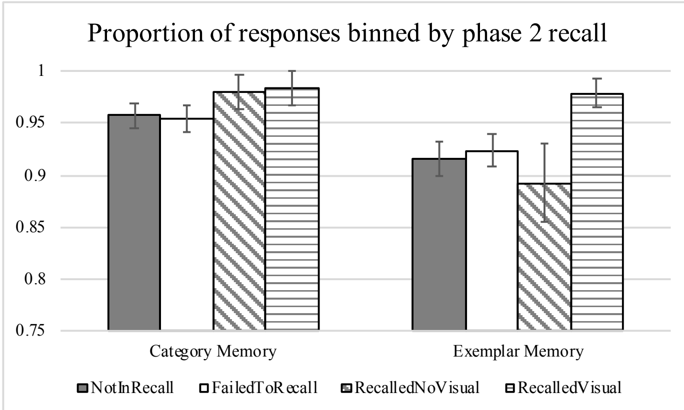

Figure S3.3. Recognition item memory results from Experiment 1, binned according to the type of recall participants did in Phase 2 (intervening recall). Category memory was the proportion of trials where the correct category was chosen (either the old object or the within-category foil) out of all trials. Exemplar memory was the proportion of trials where the old object was chosen out of the trials where the correct category was selected. Error bars show ±1S.E. of the mean.

Participants' category and exemplar memories during the final memory test were analyzed based on the nature of the intervening cued recall descriptions that were coded (Figure S3.3.). An ANOVA on type of memory (category or exemplar) and recall type (RecalledVisual, RecalledNoVisual, FailedToRecall, and NotInRecall) revealed a main effect of memory type, $F(1, 27) = 10.33$, $p < .01$, $\eta_p^2 = .28$, with better category than exemplar memory. There was a marginally significant main effect of recall type, $F(3, 81) = 2.59$, $p = .058$, $\eta_p^2 = .09$. The interaction between the two factors did not reach significance, $F(3, 81) = 1.42$, $p = .24$.

To understand how the testing effect manifests for stimuli associated with different degrees of retrieval effort, we performed planned contrasts that compared each of the three in-recall conditions with the baseline, not-in-recall condition. First, an

ANOVA comparing FailedToRecall with the NotInRecall condition, with memory type (category or exemplar) as an orthogonal factor, revealed no effect of recall condition, $F < 1$, and no interaction with memory type, $F < 1$. Thus, when cued with a scene and participants did not write down anything about an object previously presented on the scene, memory for that object did not differ from the baseline, not-in-recall condition. This is consistent with the idea that failed recall corresponds to no fruitful retrieval.

Next, we compared NotInRecall with objects for which participants recalled the category name but did not include any further visual details. An ANOVA on recall condition and memory type again revealed no main effect of recall condition, $F < 1$, and no interaction, $F < 1$.

Finally, objects in which participants recalled visual details along with the category label yielded significantly better memory than the NotInRecall baseline, $F(1, 29) = 14.60$, $p < .001$, $\eta_p^2 = .34$. This effect marginally interacted with memory type, $F(1, 29) = 3.14$, $p = .087$, $\eta_p^2 = .10$. Post-hoc paired *t*-tests showed that category memory was comparable between RecalledVisual and NotInRecall, $t(29) = 1.50$, $p = .15$, but exemplar memory was better in RecalledVisual than in NotInRecall, $t(29) = 3.55$, $p < .001$.

Discussion

Using background scene as a context cue for retrieval practice, Experiment 1 showed no overall memory enhancement of object (item) memory and category memory from recounting. This result differs from what was found in Tan and Jiang (2018), which used a category cue and found a testing effect for category memory and exemplar memory. The absence of a testing effect may have been due to a ceiling effect masking any testing benefits from recounting. Future studies are, thus, needed to determine whether there is indeed no memory facilitation from intervening testing with a more neutral context cue.

Having said that, the finer grained analysis of intervening recall does show that there is an association between the nature of recounting and later recognition, where fruitful retrieval is associated with better subsequent recognition. This was most clearly shown in exemplar memory, for objects that participants recalled with visual details. Better recognition memory was not evident when participants failed to recall an object cued by the scene. In other words, there is memory enhancement in the final recognition test for objects where a retrieval route has been established for that object. It should be noted, however, that while this positive association between intervening recounting and subsequent recognition performance is consistent with the retrieval effort account, it does not imply that there is a causal relationship between fruitful retrieval and better memory performance later on.

Experiment 2

Brewer et al. (2010) found that intervening testing of the source of items can benefit source memory later on. They found that participants were better at speaker gender discrimination if gender information was tested during the intervening test. In other words, intervening recall increased the probability of making correct source judgments by participants. As mentioned previously, the paradigm used in Experiment 1 (encoding and intervening recall) encourages retrieval of not only the item, but also the source of the item, in order to do the intervening recall task. This is especially so because the scenes were semantically unrelated to the items, preventing schema-aided guessing about the items during recall when presented the scene (Hess & Slaughter, 1990). To examine whether source memory enhancement from intervening testing can occur with this task, in the third phase of Experiment 2, we administered a recognition memory test for the scenes on which the objects were placed (i.e., the sources). If, similar to the results observed in Tan and Jiang (2018), memory for the scenes is enhanced for the objects that were included in the intervening recall test, it would grant support for the retrieval effort account and suggest that the testing effect is attribute-independent, as long as effort is exerted to retrieve the attributes during intervening testing.

Method

Participants. Thirty-two participants, 26 females and 6 males (*M* = 20.0 years, *SD* = 1.9 years) completed Experiment 2.

Materials. The same objects and scenes from Experiment 1 were also used in Experiment 2.

Design and Procedure. Experiment 2 was identical to Experiment 1, except for Phase 3 (see Figure S3.4.). On each trial in Phase 3, an object from the encoding phase was presented (image size: 4.9ºx4.9º) and the eight different scenes (each scene size: 6.3ºx6.3º) were presented side by side in a row underneath the object. The positions of the eight scenes were random. Participants were asked to click on the scene that the object had previously been presented on. A tone provided feedback.

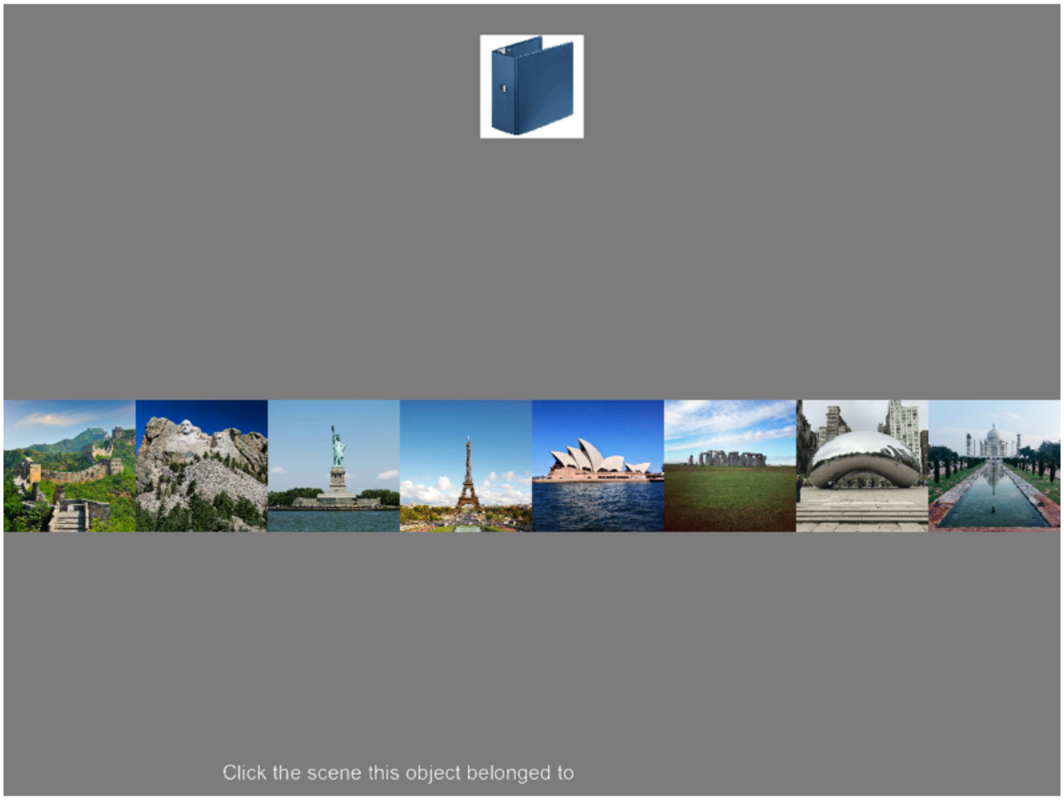

Figure S3.4. An example of the recognition phase in Experiment 2. An old object previously shown in the first encoding phase (i.e., the blue binder) was shown at the top-center part of the screen. The eight different scenes were simultaneously presented below the old object. The order of the scenes were randomized across trials. Participants were asked to click on the scene on which the object was presented previously.

Results

On each trial of the final memory test in Experiment 2, participants were shown an old object and the eight scenes that were shown in the first encoding phase. Participants' responses were either classified as incorrect (i.e., they selected the wrong scene that the object appeared on in Phase 1) or correct (i.e., they selected the scene that the object had appeared on in Phase 1).

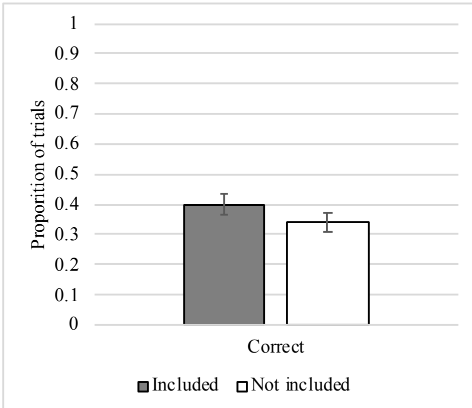

Figure S3.5. Recognition source memory results from Experiment 2 for objects included and not included in the intervening recall phase. Only correct trials are displayed. Error bars show ±1S.E. of the mean.

In the first analysis, we treated all 36 objects cued by the scenes in Phase 2 as one condition (included in recall), and compared source memory for these objects with that for objects not in recall. As shown in Figure S3.5., participants had better source memory for objects included in the intervening recall phase (*M* = 39.9%, *S.E.* = 3.6%), compared with objects not included (*M* = 34.0%, *S.E.* = 3.4%), $t(31) = 2.14$, $p < .05$.

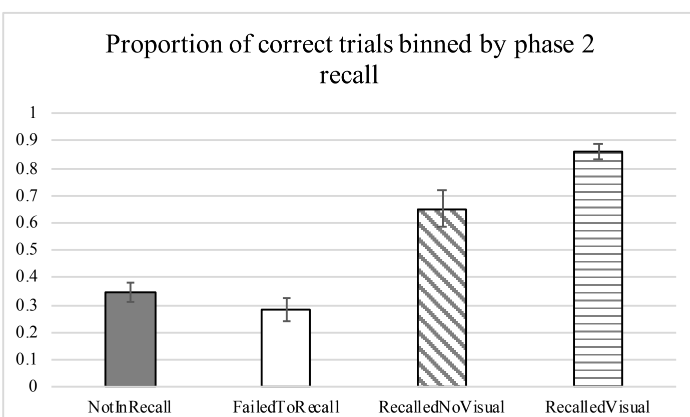

Figure S3.6. Proportion of correct trials during Phase 3 (final memory test) in Experiment 2, binned according to the type of recall participants did in Phase 2 (intervening recall). Error bars show ±1S.E. of the mean.

As in Experiment 1, we analyzed the relationship between intervening recall and subsequent recognition memory by separating trials into RecalledVisual, RecalledNoVisual, FailedToRecall, and NotInRecall, based on the descriptions provided for objects in the intervening recall phase (Figure S3.6.). Presumably, the richer the representation, and therefore, recall, of the object during intervening testing, the richer the representation of the scene-object pairing in memory as well. An ANOVA on recall type (RecalledVisual, RecalledNoVisual, FailedToRecall, and NotInRecall) for the correct scene responses showed that there was a main effect of recall type, $F(3, 57) = 34.99$, $p < .001$, $\eta_p^2 = .65$.

Further analysis showed that compared with the baseline, NotInRecall condition, when participants failed to recall the object in Phase 2, their subsequent source memory for the object was marginally worse, $t(31) = 1.87$, $p = .07$. In contrast, when participants recalled the objects, their source memory was enhanced. This was the case both for objects in which only the category label was recalled, $t(22) = 4.73$, $p < .001$, and for objects in which additional details were recalled, $t(26) = 10.70$, $p < .001$. As in Experiment 1, better recall during the intervening recall test was associated with better source memory performance in the final memory test: source memory was better for objects recalled with visual details than objects recalled without visual details, $t(19) = 2.09$, $p = .05$.

The above analysis included multiple comparisons. To avoid Type I errors associated with multiple comparisons, we also compared the three in-recall conditions, entering the levels FailedToRecall, RecalledNoVisual, and RecalledVisual, in that order. A strong linear trend was found in the source memory, $F(1, 19) = 64.91$, $p < .001$, $\eta_p^2 = .77$, supporting the idea that there is a relationship between the nature of recounting and subsequent recognition.

Discussion

In Experiment 2, intervening recall facilitated source memory. This result differed from Experiment 1 likely because source memory performance was not at ceiling, allowing for any benefits from intervening testing to be observed.

As in Experiment 1, when a more detailed analysis was conducted, an association between the nature of recounting and subsequent memory retrieval was found. Objects recalled with visual details were associated with the best source memory. Unlike Experiment 1, better recognition memory performance was observed even for objects for which no visual details were recalled (i.e., RecalledNoVisual). This discrepancy may have occurred because in Experiment 1, if the details of objects were not recalled (as in RecalledNoVisual), exemplar memory of the object was weak. In Experiment 2, however, source memory could still have been strong even when only the category of the object was recalled (RecalledNoVisual). For example, when cued with Sydney Opera House, participants may remember that they saw a suitcase but not remember any details about it (i.e., item memory was weak). Nonetheless, the connection between a suitcase and Sydney Opera House would have been retrieved. Hence, memory for the scene-object pairing could have been strong even in the RecalledNoVisual condition, eventually showing a significant positive relationship between recounting and subsequent recognition memory even for that condition in Experiment 2.

Overall, the results from Experiment 2 show that the testing effect can be observed for source memory, and not just item memory (Tan & Jiang, 2018), as long as source memory is also tested during recounting. In addition, a greater amount of retrieval of an attribute during recounting is associated with greater subsequent memory for that attribute.

General discussion

In two experiments, we examined the generalizability of intervening recall effects on visual memory. Using a neutral memory cue during intervening recall, we found that intervening recall did not enhance subsequent item exemplar memory. These results differed from that observed in Tan and Jiang (2018), which found an item memory facilitation from recounting with a category cue. The lack of memory facilitation from recounting with a context cue may have been due to a ceiling effect, preventing the observation of any testing benefits that may have resulted.

In order for participants to complete the intervening recall task, our study required that they recall both the object and the scene that it was paired with. With this design, we were able to examine the generality of intervening recall's effects on source memory. We found that unlike item memory, there was no ceiling effect for source memory, allowing for the observation of a memory facilitation from intervening recall for source memory. These results suggest that recounting visual memory can enhance not just item memory

(Tan & Jiang, 2018), but also other attributes such as source memory, if source memory is also tested. These results are consistent with previous findings (Brewer et al., 2010) which found that source memory was enhanced by intervening retrieval when source information was tested during the intervening phase. In other words, as predicted by the retrieval effort account, as long as retrieval demand and effort for the memory attribute is high enough, intervening testing can enhance subsequent memory for that attribute. Whether or not item exemplar memory can be facilitated by recounting with a more neutral context cue should be examined by future studies that eliminate the ceiling effect observed in Experiment 1.

Across both experiments, we showed that there is an association between what occurs during recounting and subsequent memory performance. More information retrieved during intervening recounting (e.g., when visual details were recalled) was related to better recognition memory performance in the final memory test for both item exemplar memory and source memory. These findings are consistent with the retrieval effort account, which states that the greater the retrieval effort exerted during intervening testing, the better the memory enhancement later on (Carpenter & DeLosh, 2006). It should be noted, however, that there is an alternative explanation for the data observed in Experiments 1 and 2. It is possible, instead, that the association shown between the nature of recounting and subsequent memory retrieval was due to certain objects and sources being better remembered in general, and thus, showing better memory in both the intervening recall phase and the final recognition memory test. Future studies are needed to investigate to what degree fruitful retrieval leads to subsequent memory enhancement. One way to examine this question is by manipulating the amount of retrieval that participants are asked to do during the intervening recall test and examining how subsequent memory performance differs across different amounts of intervening retrieval.

The generalization of the testing effect to item (Tan & Jiang, 2018) and source memory (Experiment 2) seemingly contradict past studies that have dissociated item and source memory, suggesting that they rely on different brain areas, and as a result, may have different underlying mechanisms (e.g., Glisky et al., 1995; Davachi et al., 2003; Fan et al., 2003). As mentioned previously, however, *several* regions of the brain are involved in the testing effect. Therefore, finding a dissociation, where item memory relies more on the medial temporal lobe and source memory relies more on the frontal lobe, provides little clarity as to whether intervening testing similarly affects item and source memory. Our results suggest that whether testing benefits are found for an aspect (or attribute) of memory depend more on the process (e.g., being retrieved during intervening testing) rather than the nature of the memory itself (e.g., item vs. source). Glisky et al. (1995) themselves mention that the dissociation they observed could be explained by what information was considered central versus peripheral, since the frontal lobes have been found to be involved in the encoding and retrieval of noncentral (i.e., peripheral) information (Craik & Jennings, 1992). It is possible that in their study, source information was treated as peripheral information, while item information was treated as central. This reasoning suggests that the dissociation they found may disappear if both item and source information were central to the task. By this logic, their results align with our conclusions that if retrieval demands for item and source memory were similar, intervening testing may affect them in similar ways as well.

The testing effect has been found to rely on the accuracy of intervening retrieval (Liu & Reder, 2016). In our study, we assumed that retrieval was accompanied by some level of accuracy. While we could clearly determine that there was no accurate information recalled when objects were failed to be recalled, and at least the basic-level category of the object was correct for objects that were recalled in some way, we did not analyze the accuracy of the details that were recalled. It is possible that some details recalled during the intervening phase were inaccurate. However, given that people have rich memories for the stimuli used in our study (Brady et al., 2008), it is likely that details recalled may have included accurate information. That would explain why we found that better memory in the final memory test was related to more information being retrieved for both item and source memory in the preceding recall phase. Future studies that experimentally quantify the level of accuracy during intervening retrieval are needed to more clearly understand the influence that retrieval accuracy has on the testing effect.

All in all, our results suggest that the effects of intervening testing are generalizable to different aspects (or attributes) of visual material. Having said this, however, our study is merely a first step in the examination of the generalizability of intervening testing effects across attributes. Several other questions should be investigated to increase the field's certainty about how generalizable the intervening testing effects are to all attributes of a memory. Firstly, the type of stimuli used in this study and in Tan and Jiang (2018), as mentioned before, is one that people tend to have rich memories for (Brady et al., 2008). Given that only eight scenes were used, and all were famous landmarks, it is conceivable that memory for the scenes would have been fairly good. Given the findings from past studies (Liu & Reder, 2016) that enhanced memory from intervening recall may require a certain level of accuracy, future studies should replace our stimuli with items (e.g., unfamiliar faces, see Bruce, Henderson, Newman, & Burton, 2001; Hancock, Bruce, & Burton, 2000) and sources (e.g., generic scenes) that people tend to have poorer memories for. If intervening recall impairs memory for *both* attributes, as would be expected from previous studies (Schooler & Engstler-Schooler, 1990), that would lend even more support to the conclusion that intervening recall's effects are attribute-independent.

The second question future research should investigate is whether the mechanisms underlying memory enhancement (or detriment) for both item and source memory are the same. The retrieval effort account is compatible with our findings. Besides this account, however, other theories of the testing effect have been proposed. For example, the Reconsolidation Account (Finn & Roediger, 2011), which focuses more on a re-encoding process rather than the retrieval process, suggests that after the studied material is retrieved, it becomes labile (Dudai, 2004). This enables the memory to be strengthened by re-encoding of the correctly retrieved information. There has also been evidence that both the retrieval and re-encoding processes occur, one after the other, in order to produce the testing effect (Liu, Tan, & Reder, 2018). This range of theories present the possibility that item and source memory enhancement from intervening testing may have different underlying mechanisms from each other. For instance, are both item and source memory strengthened because memory for both attributes experienced elaborative processing from increased retrieval effort, or are both types of memory strengthened by different means?

Finally, future studies should investigate whether category memory is enhanced by intervening recall. Our current study did not show the category memory enhancement from intervening recall as was observed in Tan and Jiang (2018). The lack of an effect was likely due to ceiling effects for category memory. This, in addition to the difficulty of interpreting the effects of intervening recall on category memory in Tan and Jiang (2018), due to re-exposure of the category information during intervening recall, leave open the question of how visual category memory is affected by intervening retrieval.

Our study has implications for the way in which we understand how memory works. Our findings, together with those in Tan and Jiang (2018), highlight several important factors about the effects of intervening recall on subsequent memory. One is that intervening retrieval can benefit subsequent memory as long as the memory is retrieved during intervening testing. Furthermore, across memory attributes, memory enhancement in the final memory test is related to what occurs during prior recounting. These findings suggest that in our understanding of how intervening testing affects memory, it may be more accurate and useful to examine the processes that the material (or memory) undergoes, rather than the ways in which each memory type differs from the other.

4. Section 4

Sections 2 and 3 explored the effects of recounting on one's own memory. A significant function of recounting, however, is for social communication. We are commonly asked to describe what we saw to someone else, raising the question of how well one can transmit information to another from our visual memories. Section 4 presents a set of experiments to investigate how well people can describe visual memory to others. In addition, these experiments examined how the content in participants' verbal descriptions might affect the participants' own subsequent accuracy, a question that was not investigated in the previous two sections. Judges were asked to identify the objects that participants saw, based on the participants' descriptions of the objects.

Using this procedure, we found that although people can provide useful descriptions from perception, the utility of the descriptions declines significantly with a few minutes delay between when the visual stimulus is seen and when it is described. Comparisons of participants' identification accuracy with judges' accuracy suggest that errors made by both groups stemmed from a common source. These findings highlight the danger of placing too much weight on anyone's verbal reports of what they saw and constrain theories about the extension of our memories to the external world.

Section 4 is currently under peer-review: Tan, D. H., & Jiang, Y. V. (submitted). Tell me what you saw: The usefulness of verbal descriptions for others. *Quarterly Journal of Experimental Psychology.* It is reproduced in its entirety here.

Tell me what you saw: The usefulness of verbal descriptions for others

Deborah H. Tan

Yuhong V. Jiang

Department of Psychology, University of Minnesota

Abstract

Describing what one saw to another person is common in everyday experience, such as spatial navigation and crime investigations. Past studies have examined the effects of recounting on one's own memory, neglecting an important function of memory recall in social communication. Here we report surprisingly low utility of one's verbal descriptions for others, even when visual memory for the stimuli has high capacity. Participants described photographs of common objects they had seen to enable judges to identify the target object from a foil in the same basic-level category. When describing from perception, participants were able to provide useful descriptions, allowing judges to accurately identify the target objects 87% of the time. Judges' accuracy decreased to just 57% when participants provided descriptions from memory acquired minutes ago, and to near-chance (51.8%) when the verbal descriptions were based on memory acquired 24hr ago. Comparison of participants' own identification accuracy with judges' accuracy suggests the presence of a common source of errors. The low utility of one's verbal descriptions constrains theories about the extension of one's memory to the external world and has implications for eyewitness identification and laws governing it. Our

limited ability to transmit visual information to someone else, importantly, underscores the need to corroborate verbal reports, especially in higher stakes situations, and call on more reliable sources of information whenever possible.

Keywords: visual memory, misinformation, verbal recall utility

Introduction

Extensive research has shown that recounting one's recent memory strengthens that memory and slows down forgetting. Known as the "testing effect," this finding demonstrates the beneficial effect of memory retrieval for oneself (Roediger & Karpicke, 2006). Yet in many daily activities, people recount a memory not as a mnemonic device for themselves, but to provide information to others. Delivering a lecture, retelling a story, giving directions, reporting a crime are such examples. The information provided in these recounts may guide the listener's conceptual development, spatial navigation, and crime investigation. With easy access to smartphones and the internet, another person's experience is just seconds away from being part of our own knowledge. The increasing reliance on others' memories leads us to ask: how useful are one's memory descriptions for someone else? This question has rarely been answered; the vast literature on memory focuses instead on either the accuracy of one's own memory or the accuracy of memory retrieved together with other people (e.g., Rajaram & Pereira-Pasarin, 2010). Few have examined how well one's individual memory recall can be used by someone else.

It would seem that the utility of one's memory for other people can be derived without much experimentation. For it to be used by others, the memory needs to be highly accurate and easily verbalizable. It follows that an articulate person with an accurate, verbalizable memory should be able to provide valuable information to someone else. Such communicative acts are no doubt highly effective, attested by the accumulation of a vast amount of knowledge through civilization and human history. Experimental research has therefore examined the type of memory that is less accurate or verbalizable. Concentrated around the 1980-1990's, this research examined the reliability of eyewitnesses' descriptions of crimes. Though it does not directly reveal the usefulness of the witnesses' descriptions for third-party usage, evidence suggests that the recounted information of one's visual memory may be rather sparse.

Van Koppen and Lochun (1997) examined archival data of police records from the Netherlands, containing eyewitness descriptions of robberies. The data showed that witnesses tend to provide general descriptions of the robbers such as their gender, race, and build. The witnesses gave very little specific information about the robbers, such as their facial characteristics. When their descriptions were compared with the actual robbers' characteristics, the little information that the witnesses did provide was generally accurate. This study suggests that witnesses retain some accurate memory about the crime, but their descriptions may be too sparse to be of significant value to others.

Several studies have examined the correspondence between a witness' descriptions of a perpetrator – the type of information that detectives may rely on – and the witness' own accuracy in recognizing the perpetrator from a lineup. This question was motivated by the US Supreme Court's guidelines regarding the evaluation of identification evidence (Neil v. Biggers, 1972). One guideline includes the accuracy of the witness' prior description of the perpetrator – more accurate prior description suggests that the subsequent identification through a lineup should also be more reliable. To test whether description accuracy indeed corresponds to one's own identification accuracy, Pigott and Brigham (1985) asked 120 college students to describe a person they saw for just 15 seconds. The students were later asked to identify that person from a lineup. Pigott and Brigham coded each participant's verbal description, and evaluated the correlation between the description accuracy and subsequent identification accuracy. No relationship was found – participants who provided a better description did not perform the lineup task better. The lack of correspondence is consistent with the idea that recall (e.g., the description task) and recognition (e.g., the lineup task) do not rely on the same information (Pozzulo, Lemieux, Wells, & McCuaig, 2007). It raises the possibility that a process beneficial to one's own memory – recounting an experience – may not yield useful information for others.

However, because each participant described and was tested on only one target face, the correlation between witness description and subsequent identification was examined between-participants in Pigott and Brigham (1985). Instead of suggesting a general lack of correspondence, the finding may simply indicate that good descriptors are not necessarily good identifiers (Wells, 1985).

Other studies have found some relationships between one's own identification and verbal descriptions. Wells (1985) tested 176 participants using 21 faces. Each participant viewed one face and provided a description and an identification. Wells found a modest correspondence between the two measures ($r = .27$), and attributed the correlation to item differences – more distinctive faces yield better descriptions and better identification. Other studies (Finger & Pezdek, 1999; Meissner, 2002) found that manipulating witnesses' descriptions by forcing them to provide more elaborate descriptions can adversely affect subsequent identification accuracy. Meissner (2002) showed that any inaccuracies generated during recall predicted participants' identification accuracy later on. These findings suggest that there is likely some relationship between witness description and subsequent identification. To confirm the existence of a correlation, Meissner, Sporer, and Susa (2008) conducted a meta-analysis across 33 articles. They observed a small but significant correlation, with an effect size r of .14.

The weak correspondence between descriptions from memory and identification accuracy, however, suggests that conclusions regarding the accuracy of one's memory during identification need not generalize to the descriptions produced from memory, and hence the utility of that memory to others. Unfortunately, studies reviewed above did not report how useful the witnesses' verbal descriptions were. They either did not present data on the accuracy of verbal descriptions (e.g., Pigott & Brigham, 1985), or asked raters to classify the witnesses' descriptions as "accurate", "ambiguous", or "inaccurate" (e.g., Wells, 1985). Although this rating indexes description accuracy, it is not a good measure of the utility of the descriptions. A description that is accurate but very general does not allow others to identify the specific perpetrator, and a description that is ambiguous may misguide others into choosing a foil that matches the inaccurate component of the description. A better index of the utility of a description is to examine how well other people can identify the perpetrator based on the description the witness provides.

Such an index has been used in several studies that examined different methods of description. Davies (1986) had sketch artists draw facial composites based on witnesses' verbal descriptions. Subsequently, "judges" were asked to select the target photograph from an array of photographs, based on the facial composite. There were six targets and 18 foils in the array. Judges correctly selected just 1.85 target faces on average. This finding shows that there is some, though limited, utility of verbal reports from memory to third parties. In another study, Christie and Ellis (1981) showed participants six target faces (each participant viewed only one of the six faces). After viewing a face for 60s, participants were asked to write down as much detail as they could about the face. If they did not describe any of five key features (forehead/hairstyle, eyes, nose, mouth, and chin), the experimenter prompted them to provide information about these features. The verbal descriptions were then given to judges, who relied on this information to identify the target faces from an array of six targets and 18 foils. Judges were able to identify about half of the target faces. This finding shows that when recall is carried out immediately after prolonged viewing and is guided by specific prompts, the verbal descriptions produced can be useful.

In summary, previous research has largely neglected the utility of one's memory to third-parties. Though many studies on the testing effect ask people to describe their memory, the purpose of those studies is to identify effects of retrieval effort on one's own memory (Roediger & Karpicke, 2006; Bjork, 1975), rather than the usefulness of the descriptions to others. Studies from the 1980's examined description accuracy, yet the work focused almost exclusively on eyewitnesses' descriptions of the perpetrator's face, and few studies probed the utility of those descriptions to others.

The aim of the current study is to directly examine the usefulness of one's memory descriptions to others. We extended previous work on faces to visual stimuli that are (i) more memorable, and (ii) more easily described. To this end, we asked participants to describe photographs of common objects to enable someone else to identify the described objects. The descriptions were conducted under three conditions. In Experiment 1, participants described the photographs while viewing them, eliminating memory limitation as a source of inaccuracy. In Experiment 2, participants recounted memory of the photographs they saw minutes ago. In Experiment 3, recounting was done

24 hours after exposure to the photographs. In all cases, the participants' descriptions were given to judges who had to choose the objects participants described in a two-alternative-forced-choice task (2AFC), including the described object (e.g., a tent) and a new exemplar from the same basic-level category as the target (i.e., a within-category foil) (e.g., a different tent). We examined the judges' accuracy as an index of the utility of the participants' verbal descriptions. Figure S4.1. illustrates the study procedure.

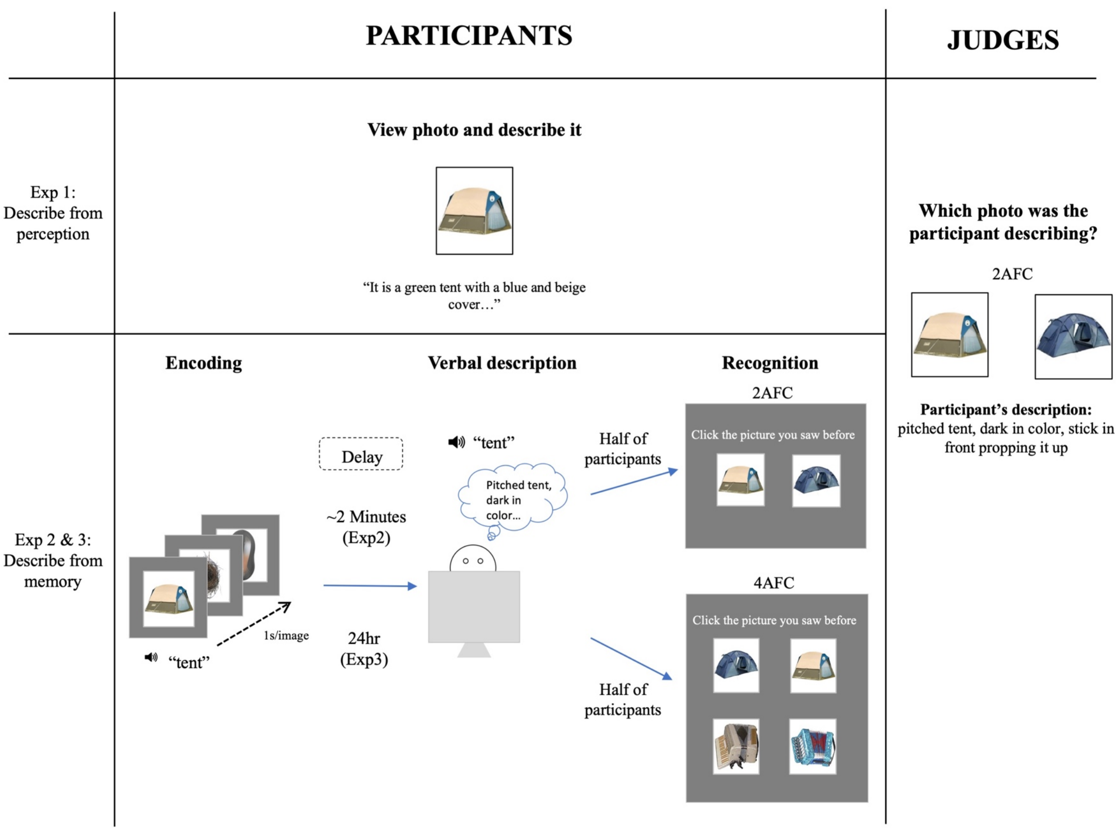

Figure S4.1. A schematic illustration of the participants' task and the judges' task. ***Participants' task (left panels):*** *In Experiment 1 (top left panel), participants were presented photographs of objects one at a time, and asked to describe each object for someone else while the object remained on the screen. In Experiments 2 and 3 (bottom left panel), participants described the photographs from memory. During the encoding* *phase, each photograph was presented at a pace of 1s/image. There was then a delay (two minutes in Experiment 2, 24hr in Experiment 3). In the next verbal description phase, participants were asked to type out a description of the named object that they saw in the preceding phase from memory. In the third recognition phase, half of the participants in each experiment were given a 2AFC task and the other half a 4AFC task, and asked to select the photograph they had seen before.* ***Judges' task (right panel):*** *A photograph of the described object and of a within-category foil were presented to judges. A participant's verbal description of the photograph they had seen was also displayed. Judges were asked to decide which photograph the participant had described.*

The study design addresses several questions. First, to what degree do previous findings from witnesses' descriptions of faces generalize to photographs of common objects? Although witnesses typically provide sparse descriptions of faces (Van Koppen & Lochun, 1997), faces are a special category of visual stimuli (Yovel & Kanwisher, 2004). Unfamiliar faces, in particular, are notoriously difficult to remember (Bruce, Henderson, Newman, & Burton, 2001; Hancock, Bruce, & Burton, 2000). In addition, face recognition is holistic, relying on the configuration of the spatial relationship between various face parts (Young, Hellawell, & Hay, 1987). Because configural information is difficult to verbalize, essential identification information of a face may not be describable. In fact, the conversion from visual to verbal codes of faces is known to impair the witnesses' *own* memory, in a phenomenon known as "verbal overshadowing" (Schooler & Engstler-Schooler, 1990). In contrast, non-face stimuli are typically recognized based on features (Young et al., 1987), making it more likely that a verbal description is useful. Critically, memory capacity for the type of common objects used in this study is extremely high. Using these same stimuli, Brady, Konkle, Alvarez, and Oliva (2008) showed that after viewing several thousand photographs, participants could identify about 90% of the objects in a 2AFC task, even when the foil was a within-category foil. They estimated that people could hold $2^{17.8}$ unique codes in visual long-term memory. The verbalizability and high-capacity raise the possibility that unlike faces, photographs of common objects may yield highly useful, transmissible memory from a witness to another person.

Second, is there any correspondence between participants' descriptions, as indexed by the judges' accuracy in choosing what participants saw, and the participants' own identification accuracy? This question pertains to the Supreme Court's guideline on the evaluation of identification evidence (Neil v. Biggers, 1972), but this time on memory of objects instead of faces. Eyewitnesses to a crime are typically asked to describe not only the perpetrator's face, but also other crime details, such as what the person was wearing and which objects were present (Davies, 1986; Pozzulo et al., 2007). The use of objects in the current study allows us to ask the same correspondence question raised previously with faces. By asking each participant to describe a large number of objects, we can evaluate the correspondence both across-participants and within-participants. In particular, do participants who provide better descriptions also recognize more objects (Pigott & Brigham, 1985)? For a given participant, are objects that have been given better descriptions also identified more accurately? Addressing these questions will reveal the utility of one's visual memory to others, shed light on the transmission of visual knowledge across individuals, and elucidate the relationship between different formats of coding (visual vs. verbal) in memory.

Experiment 1

We begin with an experiment asking participants to describe visual objects based on perception, with the objects in full view. This is a necessary first step to determine an approximate upper bound on the utility of verbal reports, when the descriptions are not limited by memory.

In Experiment 1, participants were shown a set of photographs of categorically distinct objects, one at a time. They were asked to describe each object for someone else. Typically, in eyewitness situations and everyday experience, people do not know what the foils in a line-up, for instance, could be. This makes it difficult to tune encoding and verbal descriptions to the more diagnostic features of what was seen. To mimic these real-world situations, we instructed participants to describe the objects with as much detail as possible, for the purpose of aiding someone else to identify the objects later on, without informing them of the nature of potential foils. The participants' verbal descriptions were then provided to judges, who were presented with a photograph that the participants described and a foil object that belonged to the same basic-level category as the target object. For example, the judge may be presented with a tent that participants saw and a different tent that participants did not see. The judge's task was to identify the target object based on the participant's description. We examined the judge's accuracy as an index of the utility of participants' verbal descriptions.

Even though descriptions are based on perception, several factors can limit their utility. First, information may be lost when visual representations are converted into a verbal code (Schooler & Engstler-Schooler, 1990). Second, without knowing what the foils might be, participants may fail to include diagnostic information in their descriptions. Third, judges may introduce errors when interpreting participants' verbal descriptions and identifying the objects. These factors jointly constrain the accuracy that judges can achieve, even when participants' memories are not a limiting factor.

Method

Participants. Participants in all the experiments were students from the University of Minnesota between 18-35 years of age. They were native English speakers, naïve to the purpose of the study, had normal or corrected-to-normal visual acuity, and normal color vision. Participants signed an informed consent and were compensated with extra course credit.

Twelve participants, 8 females and 4 males (*M* = 22 years, *SD* = 5.8 years), provided verbal descriptions of objects in Experiment 1.

Materials. Participants were tested individually in a room with normal interior lighting. The program was written in MATLAB (www.mathworks.com) and Psychtoolbox (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a 19" CRT monitor (1024x768 pixels) subtending 24.6ºx24.6º.

Photographs of common objects were sampled from Brady and Oliva's object database (http://cvcl.mit.edu/MM/stimuli.html). The image set contained 300 distinct basic-level categories, each with two exemplars. Each participant described 50 categorically distinct objects. The selection of the 50 categories as well as the specific exemplar shown to the participants were counterbalanced across participants.

Design and Procedure. After informed consent, participants were led to a testing room and provided with the following instructions:

"*Welcome to the study. You will be shown pictures. When the picture appears on the screen, you will also hear the name of the picture being said. Type out your description of the object. Once you are done, click "ok."…*

"Try your best to describe how the object looks like. Take your time and provide a description that will allow someone else to identify the object you saw based on your description of it.".

Participants were then shown 50 objects, one at a time, in the upper left corner of the computer monitor (image size: 14.7˚ X 14.7˚). A voice naming the object accompanied the photograph. A text window opened to the right of the photograph for participants to type out a description. Participants had unlimited time to view and describe the objects. After they had finished typing, they clicked "OK" to proceed to the next trial.

<u>Coding of the verbal descriptions.</u> Four judges, including the first author and three research assistants, completed the coding phase independently. Each judge coded all 12 participants' descriptions, one at a time. Because the photographs used as targets and foils were fully counterbalanced across participants, the judges had no information as to which photograph was more likely shown to a given participant. That is, the judges were blind to the correct answer and their coding depended entirely on the participants' verbal descriptions.

Each judge went through 50 trials of each participant's verbal descriptions. On each trial, the judge was presented with the target and a within-category foil. The two photographs were presented side-by-side, 15.7º to the left and right of fixation (each image subtended 14.7ºx14.7º). The left/right position was randomized. The verbal description the participant provided was presented below the photographs (see Figure S4.1.). The judge clicked on the image that they thought the participant saw before proceeding to the next trial.

Results

The four judges produced highly similar levels of accuracy. Mean accuracy was: judge 1 87.5% (S.E. = 3.7%), judge 2 87.8% (S.E. = 3.8%), judge 3 87.2%, (S.E. = 4.4%), and judge 4 87% (S.E. = 3.9%). The small standard error of each judge indicates that the variability in coding across the 12 participants was low. In addition, all judges produced mean accuracies within a narrow range. The mean across the four judges was 87.4% (S.E. = 0.18%). The low variability across judges suggests that their accuracy was determined mainly by the quality of the participants' descriptions, rather than by noise introduced by the judges themselves. The high consistency across judges was also observed in subsequent experiments, justifying the pooling of data across judges in the data analysis.

Using the mean data across judges as the dependent measure (Figure S4.2.), the judges' accuracy of 87.4% was significantly above chance, $t(11) = 9.57$, $p < .001$, and significantly below perfect performance, $t(11) = 3.23$, $p < .01$.

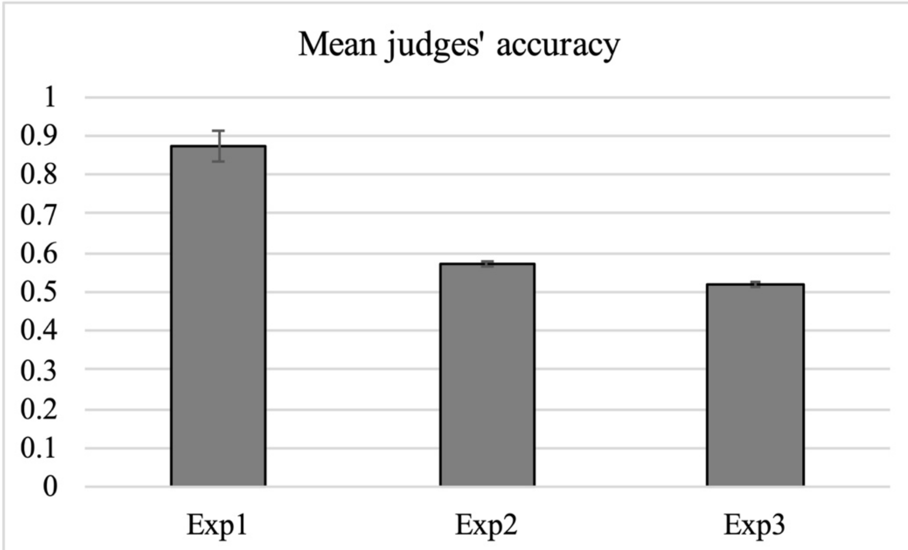

Figure S4.2. Mean judges' accuracy from Experiments 1, 2, and 3. Chance is 0.5. Error bars show ±1S.E. of the mean.

Discussion

Experiment 1 demonstrates that participants are able to provide useful verbal descriptions of visual objects, when the descriptions are given while the objects are in view. In the absence of a memory demand, participants' verbal descriptions are useful, allowing judges to discriminate the object from a within-category foil. The accuracy of the judges – around 87% - is considerably higher than previous findings using faces (Davies, 1986; Christie & Ellis, 1981). These previous studies asked people to describe faces from memory. Hence, the discrepancy between our results and theirs could be explained by a difference between perception and memory (Pigott & Brigham, 1985), and by a difference between the type of stimuli described (objects vs. faces).

The judges' accuracy is significantly below 100%, suggesting that visual information is lost when participants describe the objects. The loss of information can be attributed to two sources. First, it is not possible for participants to describe all the visual properties in words. Second, because participants did not know which features would be diagnostic in the judges' task of identifying the target object, their verbal descriptions may not have included all the diagnostic information. Both of these limitations are inherent in verbal description tasks. The 87% accuracy (rather than 100%) observed in Experiment 1 therefore serves as an approximate upper bound on what can be expected when participants describe visual objects based on memory. The next two experiments examine the degree of information loss when people describe their visual memory.

Experiment 2

Experiments 2 and 3 were part of a project investigating effects of intervening verbal recall on participants' memory of visual stimuli (Tan & Jiang, 2018). In this previous study, we showed that describing visual memory could strengthen that memory, especially when testing was done after a 24-hour delay. As in other studies on the testing effect, Tan and Jiang (2018) focused on effects of verbal description on the participants' own memory. The new focus of the current study is on the value of the verbal descriptions for third-parties who did not witness the objects.

In Experiment 2 participants provided verbal descriptions from memory acquired minutes ago. Participants engaged in an initial memory encoding phase, during which they saw a sequence of photographs one at a time. Immediately after encoding, participants were cued with the name of the photograph and asked to write down as much as they could remember about the object. As in Experiment 1, participants were asked to describe the objects for someone else's use. After the verbal description, participants' own memory for the objects was assessed using either a 4AFC or a 2AFC task. The 4AFC task presented participants with the target object they encoded, a within-category foil, and two exemplars from a new basic-level category not encoded previously (i.e., between-category foils). The 2AFC task presented participants with the target and its within-category foil. As in Experiment 1, judges were provided with participants' verbal descriptions and had to identify the object in a 2AFC task. This design allowed us to examine (i) the utility of participants' verbal descriptions in enabling the judges to choose the target object, and (ii) the correspondence between the participants' own recognition accuracy and the judges' accuracy.

Method

Data were collected as part of a project examining the testing effect on visual memory (Tan & Jiang, 2018). That study did not explore the utility of verbal descriptions for others. Here we present those data. Additional details on the testing effect can be found in Tan and Jiang (2018).

<u>Participants.</u> Forty-eight participants completed Experiment 2, which included three phases: (i) encoding, (ii) verbal description, and (iii) recognition. Half of the participants (18 females and six males, mean age 19.5 years, SD 1.6 years) were tested in a 4AFC recognition task in Phase 3, whereas the other half (13 females and 11 males, mean age 19.8 years, SD 1.7 years) were tested in a 2AFC recognition task.

<u>Materials.</u> Objects were drawn from the same object database as in Experiment 1, plus another 10 basic-level categories of objects that served as fillers in the encoding phase (see procedure below).

Phase 1: encoding. Participants first engaged in an encoding phase. During this phase, participants were presented with a sequence of visual objects at a pace of 1s/image. Each object was presented at the center of the screen (24.6ºx24.6º). Participants were asked to remember the objects. To ensure task compliance, participants were asked to monitor occasional repetitions of objects. The encoding sequence had 150 objects that were presented once, and 10 filler objects that were presented twice at random intervals. The fillers were not included in subsequent phases. For 12 participants, the objects were presented on their own, and participants pressed the spacebar whenever they detected an object that appeared earlier in the sequence. For the other 36 participants, a voice naming the object was presented along with its visual presentation. The voice tied up the computer's processor, preventing it from collecting button presses. These participants reported the number of repetitions they detected at the end of the encoding phase. Repetition detection accuracy was high (as reported in Tan & Jiang, 2018), suggesting that participants had complied with the task instruction. All the objects presented in the encoding phase were unique in terms of their basic-level categories. A random half of the 300 categories was assigned for encoding, while the other half was assigned as foils for the recognition phase (Phase 3). The assignment was reversed for half of the participants. Object categories and exemplars used for targets and foils were fully counterbalanced across participants.

Phase 2: verbal description. A random third of the encoded objects ($N = 50$) were used in this second phase. Assignment of stimuli for inclusion in Phase 2 was counterbalanced across participants. Participants were shown the written name of an object on the screen, and a voice announcing the object's name was played. Participants were asked to type out a detailed description of the named object that they had seen in Phase 1 to enable someone else to identify the object based on their description. After they had finished typing, they clicked “OK” to proceed to the next trial. No feedback was given.

Phase 3: recognition memory test. The recognition memory test included 150 trials, including the 50 objects that participants were cued to recall in Phase 2, and the 100 objects not included in Phase 2. Our focus in this study is on participants’ recognition memory for the 50 objects they described in Phase 2. Memory for the other 100 objects was used to probe the testing effect, and data were reported previously (Tan & Jiang, 2018).

For a subset of the participants (*N* = 24), on each trial, four objects were presented in a 4AFC task, one in each quadrant (image size: 14.7˚ X 14.7˚; the center of each image was 11.7˚ from fixation, horizontally displaced by 7.9˚ and vertically displaced by 8.6˚ from fixation). One object was the target object, which participants had seen in Phase 1. Another was a within-category foil, and the other two were between-category foils. None of the foils were previously presented. The positions of the four objects were random. Participants were asked to click on the object they had seen before. A tone provided feedback. For the other subset of participants (*N* = 24), only two options (2AFC) were presented side by side on the horizontal meridian (the center of each image was 7.9˚ away from fixation). One option was the target, while the other was the within-category foil. Participants were asked to click on the object they had seen before and a tone provided feedback.

Coding of the verbal descriptions. Three research assistants served as judges. As in Experiment 1, the judges were blind to the images that each participant saw. The coding procedure was the same as in Experiment 1. The judges were presented with each participant's verbal description, along with the target object and a within-category foil. Judges had to identify the target object.

Results

1. Judges' coding accuracy

Judges' accuracy in identifying the target objects was surprisingly low (Figure S4.2.). Mean accuracy was: judge 1 55.6% (S.E. = 1.4%), judge 2 58.3% (S.E. = 1.03%), and judge 3 57.4% (S.E. = 1.02%). As in Experiment 1, for each judge, accuracy across the 48 participants they coded fell within a narrow range. In addition, the three judges produced accuracies within 2.7% of each other. The mean for all judges was 57.1% (S.E. = 0.86%). The consistency across judges justified the pooling of data across all judges to provide an index of the utility of participants' descriptions to others.

We compared the judges' accuracy for the 12 participants in Experiment 1, with judges' accuracy for the 48 participants in Experiment 2. This revealed a significant difference, $t(58) = 11.8$, $p < .001$, suggesting that the utility of verbal descriptions substantially declined when it was based on memory rather than perception. The judges' accuracy in Experiment 2 was above chance (50%), $t(47) = 8.27$, $p < .001$, suggesting the presence of some, though limited, utility of one's memory descriptions to third parties.

2. Difference between judges' accuracy and participants' own identification

Participants' own identification exceeded the judges' accuracy. Participants tested in the 2AFC recognition task essentially carried out the same task as the judges, except that their choice was based on their own memory rather than on the descriptions they provided. For the 50 objects they described, the 24 participants in the 2AFC recognition task had a mean accuracy of 67.7% (S.E. = 2.1%), significantly higher than the judges' accuracy, 56.9% (S.E. = 1.3%), $t(23) = 7.29$, $p < .001$. Participants tested in the 4AFC task identified the correct category (either the target or the within-category foil) 88.1% (S.E. = 2.7%) of the time, suggesting good category memory. Among trials in which they identified the correct category, participants chose the target 70.4% (S.E. = 2.3%) of the time. This value was again significantly higher than the judges' accuracy (M = 57.3%, *S.E.* = 1.2%), $t(23) = 7.37$, $p < .001$. The difference between participants' own recognition accuracy and the judges' identification accuracy indicates information loss when visual memory was transmitted to another person via verbal description. The loss could be due to a change in coding format (from visual to verbal) or to the failure of fully describing one's visual memory.

3. <u>Correspondence between judges' accuracy and participants' own identification</u>

Although judges performed worse than the participants, the pattern of behavior across images showed a clear correspondence between the two (Figure S4.3.).

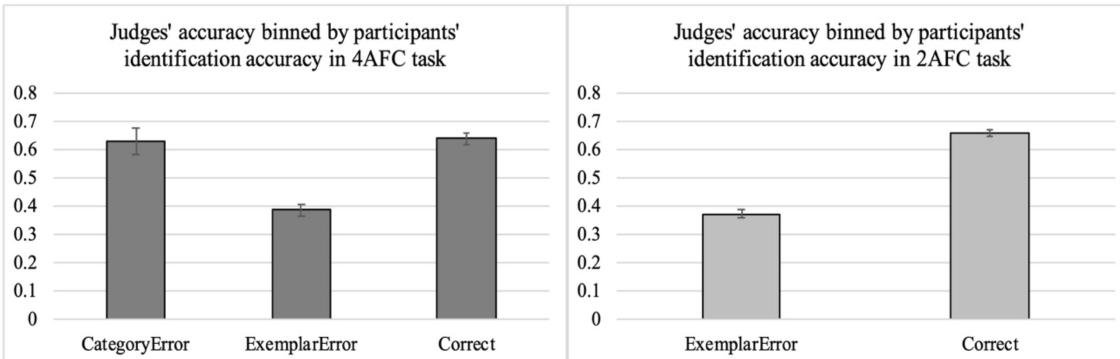

Figure S4.3. Judges' mean accuracy binned by participants' identification accuracy in Phase 3 (recognition memory test). Left: Results when participants were given the 4AFC task. Participants either committed a category error (i.e., choosing a between-category foil), an exemplar error (i.e., choosing a within-category foil), or chose the target and obtained a correct response. Right: Results when participants were given the 2AFC task. Error bars show ±1S.E. of the mean.

When participants were given the 4AFC task in Phase 3 (Figure S4.3., left), judges were significantly above chance ($M = 63.9\%$, $S.E. = 2.0\%$) for images that participants themselves successfully identified later on, $t(23) = 6.99$, $p < .001$. Judges' accuracy declined to significantly below chance ($M = 38.7\%$, $S.E. = 2.1\%$) for images that participants later chose the within-category foil, $t(23) = 5.46$, $p < .001$. This suggests that when participants were mistaken with regard to which exemplar they had seen, their earlier verbal description was more consistent with the foil, misleading the judges toward choosing the foil more often than the target. Finally, for images that participants later chose a between-category foil, judges' accuracy was above chance (*M* = 62.7%, *S.E.* = 4.7%), $t(19) = 2.72$, $p = .013$[1].

Correspondence between the participants' own identification and the judges' choices was also observed in participants who completed the 2AFC task in Phase 3 (Figure S4.3., right). Here, judges performed significantly above chance (*M* = 66%, *S.E.* = 1.3%) when participants later chose the target, $t(23) = 11.78$, $p < .001$, and performed significantly below chance (*M* = 37%, *S.E.* = 1.6%) when participants later chose the within-category foil, $t(23) = 7.93$, $p < .001$.

4. <u>Good vs. bad descriptors, good vs. bad identifiers</u>

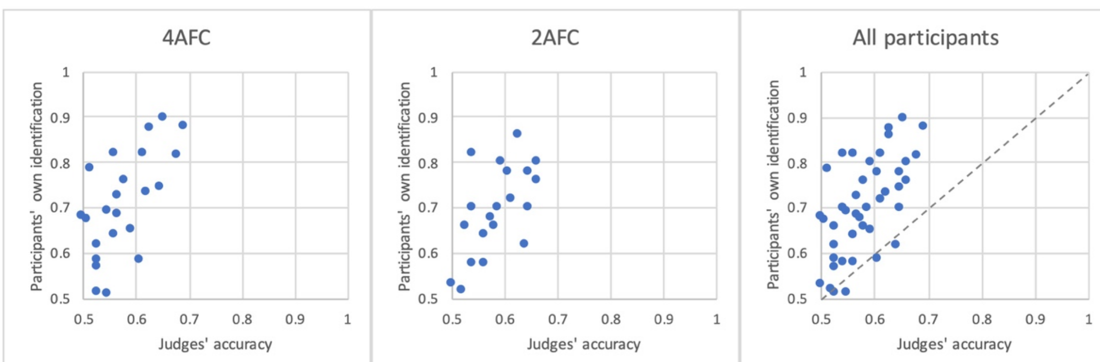

Figure S4.4. Scatterplots on the relationship between the judges' accuracy and the participants' own identification accuracy in Experiment 2. Each data point corresponds to a single participant.

[1] The *N* was smaller for between-category foils because not every participant made between-category errors.

The correspondence examined above was across stimuli – we found that objects that participants correctly identified were also more accurately described, compared with objects that participants misidentified. Here we investigated the between-participant variability in description and identification. Each participant yielded an identification score – percent of the time that they chose the target rather than the within-category foil. Each participant also yielded a description score – percent of the time that the judges chose the target based on their verbal descriptions. Figure S4.4. plots the correlation between the two measures, for participants in the 4AFC task (Figure S4.4., left), 2AFC task (Figure S4.4., middle), and for all participants (Figure S4.4., right). We found a significant correlation between the two measures, with Pearson's correlation coefficient of .65 in the 4AFC task ($p = .001$), .70 in the 2AFC task ($p < .001$), and .67 when all participants were considered ($p < .001$). Consistent with the previous meta-analysis (Meissner et al., 2008), participants who provided better descriptions were also better identifiers themselves. Information loss is also apparent in Figure S4.4. (right), where the main diagonal shows equivalent performance between the judges and the participants. Nearly all participants' data lie above the diagonal line, meaning that participants' own identification accuracy exceeds the judges' accuracy.

Discussion

Experiment 2 shows that when describing objects from memory, participants provide some, though highly limited, information to others. Judges can only discriminate the target object participants saw from a within-category exemplar 57% of the time. This level is about 30% lower than that observed in Experiment 1 in which descriptions were based on perception rather than memory. Participants' own identification accuracy exceeds the judges', showing information loss when describing their visual memory. Thus, the judges' poor identification accuracy can be attributed to participants' memory inaccuracy, as well as to information loss when visual memory is transcribed into words.

This experiment also reveals strong correspondence between participants' own identification accuracy and the utility of their descriptions. Participants who give better descriptions and therefore enable judges to reach higher accuracy, also themselves have higher identification accuracy. In addition, within a participant, objects receiving better descriptions are also ones that are more accurately identified by the participant. When participants choose the wrong exemplar, their earlier verbal description contains misinformation, resulting in judges' tendency to choose the wrong exemplar more often than the target object.

The correspondence between the participants' description accuracy and their identification accuracy appears to be stronger in the current study than in previous studies (e.g., Pigott & Brigham, 1985). This may be due to stimulus differences: faces may be so difficult to describe such that what is described has poor correspondence to the actual stimuli. Another difference is the way description accuracy is measured. Several previous studies measured description accuracy by rating the descriptions as "accurate", "ambiguous", or "inaccurate" (Wells, 1985). This classification may inflate the utility of the descriptions, given that an accurate but general description may not be useful, and a partly accurate description may be misleading. In addition, the ratings used in previous studies may lack reliability. In a pilot phase of our study, we asked raters to rate participants' verbal descriptions by displaying the verbal descriptions along with the actual object they saw. We found that the raters lacked confidence in rating the descriptions – many statements contained both correct and misleading information and could not be easily classified as "partly accurate" or "inaccurate." Instead of forcing the raters to make their best bet on the classification, the current study uses a method that is unambiguous – judges are asked to identify the target object based on the participants' verbal descriptions. This way, everything in the verbal descriptions becomes relevant, and the judges' task is essentially the same as the participants' identification task (Phase 3). The low variability across judges indicates that this method is highly reliable. The strong correspondence between the participants' identification accuracy and the utility of their descriptions provides support for one of the Supreme Court's guidelines regarding the evaluation of identification evidence (Neil v. Biggers, 1972). Accurate earlier descriptions indeed have some correspondence to accurate subsequent identification, at least for visual memory of objects.

Experiment 3

Recounting one's memory can be fairly immediate, as in Experiment 2. Often times, however, there is a delay before we can provide a verbal description of what we saw. How does the utility of verbal reports change with time? The retrieval effort account (Carpenter & DeLosh, 2006; Carpenter, 2009) states that people remember what was encoded better if the time interval between the initial study and the first retrieval is longer. This suggests that delaying the first retrieval attempt makes the solution less accessible, necessitating an increase in retrieval effort on that attempt and enhancing one's memory in subsequent tests. Along these lines, it is possible that a longer delay between encoding and the verbal recall in our study would lead to a more effortful retrieval during verbal recall, yielding a verbal description that has greater utility to third parties (e.g., perhaps only the more diagnostic information is retained in that description). On the other hand, memory decays during delays, and this may decrease the verbal description utility to others. As shown in Experiment 2, verbal descriptions from memory are much less useful than those from perception, suggesting that as memory further decays, the utility of the verbal descriptions may decline further as well. Experiment 3 tests these competing predictions by inserting a 24-hour delay between encoding and the verbal description phase.

Method

Participants. Forty-five participants, 33 females and 12 males, completed the study. Their mean age was 19 years.

Design and Procedure. Experiment 3 was identical to Experiment 2, except that a delay of 24 hours was inserted between Phases 1 and 2. Participants encoded 150 photographs on Day 1. They returned 24 hours later and completed a verbal description of 50 objects in a similar manner as in Experiment 2. Immediately after the verbal description, participants completed the identification task using either the 4AFC (N=23) or the 2AFC task (N=22).

Coding of the verbal descriptions. Three research assistants served as judges in the coding phase. The procedure was the same as in Experiments 1 and 2.

Results

1. <u>Judges' coding accuracy</u>

Judges performed poorly based on participants' verbal descriptions of the objects (Figure S4.2.). The mean accuracy was: judge 1 51.3% (S.E. = 0.97%), judge 2 51.5% (S.E. = 1.0%), and judge 3 52.7% (S.E. = 1.2%). As in the first two experiments, the judges produced highly similar levels of accuracy, with a mean of 51.8% and an S.E. of 0.8%. The judges' data were therefore averaged to produce an index of the utility of the participants' verbal descriptions.

Judges' accuracy – 51.8% – was very low. It was statistically above chance, $t(44) = 2.25$, $p = .03$, but significantly lower than that observed in Experiment 2, $t(91) = 4.46$, $p < .001$, or Experiment 1, $t(55) = 14.10$, $p < .001$. Thus, following a 24-hour delay, the verbal descriptions were of little value to third parties.

2. <u>Difference between judges' accuracy and participants' own identification</u>

As in Experiment 2, the judges' accuracy was lower than participants' own identification accuracy. In Phase 3, for the 50 objects they described, participants tested in the 4AFC task rarely chose the wrong category (correct category was chosen 87.2% of the time, S.E. = 1.7%). Among trials in which the correct object category was chosen, participants chose the target object (as opposed to the within-category foil) 63.4% of the time (S.E. = 1.4%). This was significantly higher than the judges' accuracy ($M = 52.3\%$, *S.E.* = 1.4%), $t(22) = 8.86$, $p < .001$. Participants tested in the 2AFC task chose the target object 63% of the time (S.E. = 1.6%), again significantly higher than the judges' accuracy

(M = 51.3%, $S.E.$ = 0.87%), $t(21) = 7.58$, $p < .001$. Clearly, participants had access to information that was not part of their verbal descriptions.

3. <u>Correspondence between judges' accuracy and participants' own identification</u>

Even though the judges' overall accuracy was close to chance, they did not perform randomly. Instead, their performance corresponded strongly with the participants' performance. First, when participants were given the 4AFC task in Phase 3, judges were significantly above chance (M = 60%, $S.E.$ = 2.2%) for images that participants themselves successfully identified later, $t(22) = 4.57$, $p < .001$. Judges' accuracy declined to a level not different from chance (M = 47%, $S.E.$ = 4.2%) when participants later chose a between-category foil, $t(21) = 0.66$, $p = .52$[2]. Judges were significantly below chance (M = 41%. $S.E.$ = 2.1%) when participants later chose the within-category foil instead of the target, $t(22) = 4.18$, $p < .001$. Similar results were found in participants who completed the 2AFC recognition task. For images that participants later accurately identified, the judges also performed above chance based on the participants' descriptions (M = 59%, $S.E.$ = 1.6%), $t(21) = 5.83$, $p < .001$. For images that participants later incorrectly identified, judges were also more likely to choose the within-category foil – the judges' accuracy was 36% (S.E. = 2.0%), a level significantly below chance, $t(21) = 6.78$, $p < .001$. These data confirm once again that participants' verbal descriptions contained both accurate and misleading information and that this information was the common source of errors for both the participants themselves and for the judges.

---

[2] The degrees of freedom was 21 rather than 22 because one participant did not commit any between-category errors.

4. <u>Good vs. bad descriptors, good vs. bad identifiers</u>

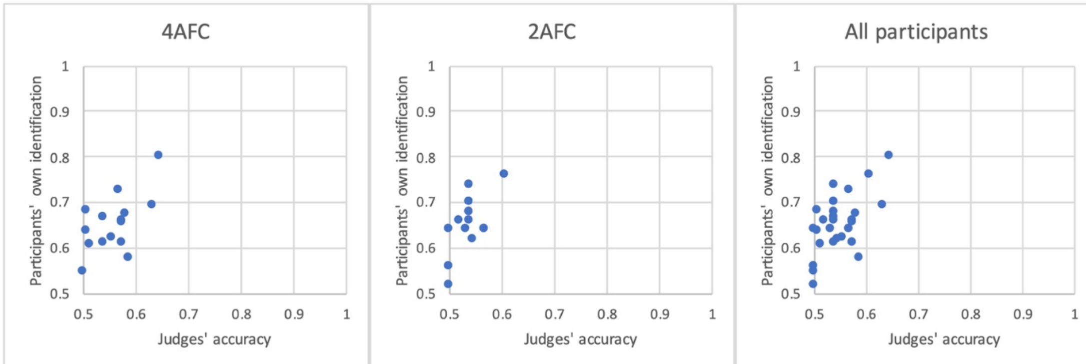

Figure S4.5. Scatterplots on the relationship between the judges' accuracy and the participants' own identification accuracy in Experiment 3. Each data point corresponds to a single participant.

As in Experiment 2, for each participant, we computed their average identification accuracy and the judges' accuracy based on their descriptions (Figure S4.5.). The results suggest that good descriptors also tended to be good identifiers. For participants who completed the 4AFC task (Figure S4.5., left), Pearson's correlation between the participants' and the judges' accuracy was .48, p = .02. The correlation failed to reach significance in participants who completed the 2AFC task (Figure S4.5., middle), Pearson's r = 0.37, p = .09. When all participants were combined (Figure S4.5., right), the correlation was r = 0.43, p = .004. These correlations were lower than in Experiment 2, possibly due to near-floor performance of the judges, restricting the range of data and lowering the correlations.

Discussion

With a 24-hour delay before verbal recall, the utility of verbal reports decreases to a level near chance. This finding suggests that time delay causes memory for the visual stimuli to decay such that people are less able to provide useful verbal descriptions for third parties. Even though people are suggested to have excellent memories for the visual details of common objects (Brady et al., 2008), this memory does not necessarily translate to useful descriptions for third parties' use, as compared to when the objects are described from perception.

Despite very low utility for third parties, Experiment 3 again confirmed strong correspondence between participants' identification based on their visual memory, and the judges' choices based on participants' verbal descriptions. When participants were correct in identification, their earlier verbal descriptions contained sufficiently accurate information for the judges to perform above chance. When participants were incorrect in identification, their earlier reports contained inaccurate information that misled the judges toward choosing the wrong object. This finding suggests that verbal recall and visual identification share common memory sources (Meissner, 2002).

General Discussion

In this study we examined how useful verbal reports of objects are for third parties. We found that when verbal descriptions are constructed from perception, people are able to provide very useful descriptions. However, when verbal descriptions are provided from memory, utility of the descriptions drops drastically. Minutes after the initial memory encoding, the verbal descriptions enable judges to choose the target object with just 57% accuracy, where chance is 50%. Twenty-four hours later, the utility declines to 51.8%, a level just barely above chance. The participants' ability to produce a significantly more useful verbal report of the objects from perception compared to memory suggests that the low verbal report utility is largely due to memory limitations, rather than poor verbal descriptive ability overall.

The contrast between perception and memory sheds light on the debate between two camps of thought on visual memory. One camp states that people have highly impressive visual long-term memory. For example, Standing, Conezio, and Haber (1970) showed that people can recognize more than 90% of photographs presented a day earlier, even though they initially saw each of several thousand photographs for just 10s. Brady et al. (2008) found that people have accurate memory about the state and exemplar of several thousand visual objects they viewed for just 3s each. They estimate that visual long-term memory can hold $2^{17.8}$ bits of unique information. In contrast, others propose that instead of representing visual details in memory, people rely on the external world as their "memory" (O'Regan & Noe, 2001), and visual details are lost as soon as they are no longer in view (Levin & Simons, 1997). In a virtual reality block-construction task, Ballard, Hayhoe, and Pelz (1995) found that people tend to code just one piece of information (e.g., the color of a block) at a time, and frequently make repeated fixations on the same object to extract additional information (e.g., the shape of a block). So, does visual memory have high capacity or is it sparse? There is no doubt these contrasting views can both be correct under different circumstances. The question is, which view more accurately captures performance in specific situations? Our study shows that when it comes to the transmission of visual information from one person to another, memory is a highly unreliable source. The idealized, high-capacity visual memory people are able to achieve (Brady et al., 2008) does not lend itself to a useful verbal report for others. When allowed to look at the visual stimulus itself, however, people do notice details and are able to describe them to produce a useful report for third parties.

Although memory is a significant source of inaccuracies found in this study, we also have some evidence that information is lost when people translate visual codes to verbal codes. In Experiment 1 when participants described objects from perception, judges were able to identify the target object at 87%, a level significantly below perfect. In Experiments 2 and 3 when participants described objects from memory, the judges' accuracy was significantly below the participants' own identification accuracy. This loss of information may be due to a variety of reasons. Part of the information loss may stem from a mismatch between the non-linguistic processing that the recognition task involves and the linguistic verbal processing that the recall task involves (Pozzulo et al., 2007). Additionally, with perception-based descriptions, participants could rely on the picture to cue them to features that could be included in their descriptions. With memory-based descriptions however, a less precise representation of the object, which fades with time, may make the translation more difficult. Finally, because participants had no prior knowledge of the nature of the foils, they could not tune their encoding and descriptions to the diagnostic aspects of the objects. This means that even if they had access to diagnostic features, they may not have included that information in the verbal descriptions. For these reasons, people are likely to have difficulty representing the full extent of their visual memory in verbal reports.

Even so, it is clear that some information is still transmitted to the judges. In both Experiments 2 and 3, there was a systematic pattern in the judges' performance which aligned with how participants themselves performed in the final memory test. Judges performed better (i.e., participants' verbal descriptions were more useful) when participants themselves recognized those objects later on. Conversely, when participants later chose the within-category foil, the verbal descriptions they provided for those objects also misled judges to choose the within-category foil more often than the target. These findings show that verbal descriptions can contain useful, as well as misleading, information for third parties, and that the misinformation was a source of the judges' inaccurate choices. Unfortunately, the misinformation may also have been a source of the participants' own mistakes. When participants attempted to recall the objects in Phase 2, they may have had inaccurate memories to begin with, leading to incorrect recognition in Phase 3. Even if their memory did not contain misleading information prior to recall, the act of recalling the objects may have introduced misinformation (Brainerd & Reyna, 1998a; Reyna, 1995), which could then get incorporated into memory through reconsolidation (Chan, Thomas, & Bulevich, 2009). In this study, objects included in Phase 2 may be associated with worse memory than objects omitted from verbal recall, and there is some evidence that this is the case (Tan & Jiang, 2018).

Although recall and recognition may rely on different sources of information (Pozzulo et al., 2007), the strong correspondence between the participants' identification and the judges' choice suggests that they do share common sources. This finding is consistent with previous meta-analysis on eyewitness testimony (Meissner et al., 2008). The correlation observed in our study appears to be stronger than what was found in the past. This may be due to the use of objects, which are more verbalizable, rather than faces in the present study, and the similarity between the judges' task and the participants'. Instead of asking the judges to rate the participants' verbal descriptions as accurate or inaccurate, our study design involved the same identification task for the judges and the participants, changing only the basis of the identification (i.e., someone else's verbal description vs. one's own experience). Our study supports the Supreme Court's guidelines with regards to evaluating identification evidence (Neil v. Biggers, 1972): better prior description of crime details does correspond to more accurate identification subsequently.

Intuitively, both the reduction in information from perception to memory and the loss of information from visual to verbal codes are expected. Nevertheless, it is surprising that these factors together nearly abolish the utility of one's verbal descriptions to others. Even for a type of stimuli that is somewhat verbalizable and associated with high memory capacity, the judges' identification accuracy was close to chance after just minutes (or 24 hours) of delay. This finding presents strong constraints to the utility of one's visual memory. As we gain ready access to other people's knowledge, via text messaging or other social media, there is no guarantee that the expansion of cognition from one's own brain to the external world would bring the type of gains we would like to have. The unreliability of information transmission from one person to another suggests, instead, that such transmission can be a major source of errors, a hotbed for "fake news." This is not to say that accurate information cannot be transmitted. Though visual memory for exemplars within a category may not be accurately transmitted, visual memory for categorical information may be better preserved. In addition, conditions that improve the initial memory, such as increased encoding duration, repetition, and a reduction in the number of objects encoded, may all be beneficial. Nonetheless, the situation presented here – brief exposure to multiple objects and immediate or delayed recall – is typical of daily experience, including eyewitness situations.

Our findings suggest that in real-world situations that resemble the current study, it would be more beneficial to depend on surveillance images or video records rather than witnesses' verbal descriptions when trying to decipher the visual details of what was seen previously. The low level of accuracy observed in this study is dangerous in higher stakes situations, such as during the identification of a critical object used in a crime. Our study only presented two choices, while in reality, an infinite number of possibilities for an object often exist, likely lowering "judges'" accuracy further. The strong correspondence between the judges' choice and the participants' own identification, even in cases when both were wrong, suggests that internal consistency is no proof for accuracy. Unless corroborated by additional sources, such internal consistency may further misguide investigations, allowing inaccurate information to perpetuate through multiple channels. In cases where surveillance records are unavailable, a more non-linguistic task, such as pointing out and describing similarities and differences (Brown & Lloyd-Jones, 2005) between the person's memory and another exemplar presented visually, rather than a pure verbal report task, should be used to uncover people's visual memories. Relying solely on verbal reports likely prevents significant portions of a person's visual memory from being transmitted to third parties, as observed from the discrepancy between the judges' accuracy and the participants' own identification accuracy.

The advancement of technologies, including search engines and social media, has significantly changed the way people access information. Because the external world often functions as an extension of one's own cognition (O'Regan & Noë, 2001), as our access to the external world changes, it also profoundly affects how we function. Instead of remembering essential information, for instance, people now tend to remember where to find it (Sparrow, Liu, & Wegner, 2011). As people gain instant access to others' knowledge through social media and other apps, we become increasingly reliant on others to guide our actions, plans, and understanding. In this study, we show that there are constraints on the reliability and utility of others' knowledge to us. Other people's visual memories, in particular, may be useless or even misleading, after just a day of delay. This limitation underscores the importance of more reliable sources of information, and may help us understand the perpetuation of misinformation in the modern world.

5. Summary and implications

The goal of this dissertation was to examine the consequences of recounting a memory for both the describer and the listener. To this end, three sets of experiments were conducted, two of which investigated the effects of recounting on a describer's subsequent memory, and one of which examined the utility of that recounting for third parties. In this final section, I will summarize the findings from the studies done and discuss their implications for both the scientific field and for society.

5.1. The effects of recounting on visual memory

<u>5.1.1. How recounting affects visual exemplar memory - findings</u>

This first study examined the effects of intervening recall on visual exemplar (i.e., detail) memory for common objects. Participants were shown photographs of common objects one at a time, and then asked to verbally recall a subset of them. In a final memory test, recognition memory for all the objects was tested. We found that intervening recall had multiple effects on visual memory, showing that recounting introduces interference in the short term, but produces a net positive effect on visual memory in the long term. Intervening recall slowed forgetting of the materials such that after a 24-hour delay following intervening recall, photographs that were recalled were better remembered than those that were not recalled.

<u>5.1.2. The effects of recounting across memory attributes - findings</u>

In the second study, I extended the investigation of how generalizable the effects of recounting a memory are on subsequent memory performance. Specifically, I ask whether the effects of recounting generalize across different attributes of memory – namely, item and source memory. Participants were shown photographs of common objects superimposed onto scenes. In a subsequent intervening recall phase, they were shown a subset of the scenes and asked to verbally recall all the objects that were previously presented on that scene. This task required that both the objects and their corresponding scene be recalled. Hence, retrieval effort was needed for both item and source memory during this intervening phase. In a final memory test 24 hours later, participants' memory for either the objects (item) or the scenes (source) were tested. The item exemplar memory facilitation previously observed from intervening recall in Section 2 was extended to source memory. No memory enhancement from recounting was observed for item exemplar memory in this second study, although this could have been due to the presence of a ceiling effect. Analysis of the participants' verbal descriptions suggested that the intervening recall benefits were related to how much fruitful retrieval occurred during recall, lending support to the retrieval effort account (Carpenter & DeLosh, 2006) of the testing effect. Whether a causal relationship between the amount of retrieval and subsequent memory performance exists should be examined by future studies.

5.1.3. Significance of the findings

5.1.3.1. Evidence for verbal overshadowing and spontaneous misinformation

The findings from the first two studies shed light on several controversial theories surrounding the effects of recounting, the first of which is verbal overshadowing (Schooler & Engstler-Schooler, 1990). The verbal overshadowing effect is said to occur because of transfer inappropriate retrieval (Schooler, Fiore, & Brandimonte, 1997). A visual stimulus (often a perpetrator's face) contains both verbalizable and less verbalizable components. By attempting to verbally describe the face, the process overshadows the components that the witness is not able to verbalize. This suppression then leads to increased difficulty while trying to match the now description-based memory of the face to the visual stimuli presented during the recognition task.

Particularly in the 1990's and early 2000's, after the effect was first reported (Schooler & Engstler-Schooler, 1990), the verbal overshadowing effect was argued to either be difficult to replicate (Yu & Geiselman, 1993) or smaller than had originally been reported (Meissner & Brigham, 2001). These findings raised doubts about whether the phenomenon existed at all. Given the variation in methods used across studies, a large-scale replication of the original Schooler & Engstler-Schooler (1990) study was done (Alogna et al., 2014). It was found that the effect largely depended on when the tasks (i.e., witnessing of the event, describing the event, and performing the recognition task after describing) were performed relative to each other. The verbal overshadowing effect was larger when a delay was introduced after the event was witnessed and before the verbal description was provided. The suggestion that the timing of the tasks affects the magnitude of the effect is also supported by the results from a meta-analysis (Meissner & Brigham, 2001), which found that the verbal overshadowing effect was evident when the final recognition task occurred immediately after the verbal description was provided. When a delay was inserted between the verbal description task and the final recognition task, however, the verbal overshadowing effect disappeared. This may have been due to a 'release' of the interference from verbalization over the delay.

The results in Section 2 align with these findings and suggest that the verbal overshadowing effect can indeed be observed and does exist, given the “right” conditions. Schooler et al. (1997) suggest that it is the suppression of the less verbalizable components of memory, from verbalization, that leads to memory performance impairment. In Section 2, where category memory and exemplar memory were separately analyzed, exemplar memory was, by definition, the less verbalizable component of participants’ memories, and thus should have been most adversely affected by verbal recall. Indeed, that was the case. When final memory testing occurred immediately after intervening recall, an interference effect was observed only for exemplar memory. However, once a delay was inserted between intervening recall and final memory test, the impairment was no longer observed, similar to that found in past studies (Meissner & Brigham, 2001). In fact, a net positive effect emerged. Further evidence of the verbal overshadowing effect comes from the data observed in Section 2 Experiment 2, where asking participants to describe (and not just visualize) the object during intervening recall led to the largest impairment in exemplar memory. These findings suggest that the act of verbalizing one’s memory for the visual objects does play a role in the interference effect that was observed in experiments where the final memory test occurred immediately after intervening recall. The findings in Section 2, therefore, strengthen the idea that the verbal overshadowing effect can be one reason why intervening recall leads to subsequent memory impairment.

Another way in which memory can be interfered with is through reactivation and reconsolidation (Dudai, 2004; Diekelmann et al., 2011; Schiller et al., 2010). Whether or not the memory interference observed in past research is actually due to reconsolidation effects, however, is still debated (Elsey et al., 2018). What is clear is that memory can be interfered with when retrieved. Studies that have examined the role of reconsolidation in memory interference have typically involved experimenter-induced misinformation where the experimenter observes whether the misinformation they provide gets incorporated into and interferes with the participants' original memory (e.g., Chan & LaPaglia, 2013; Chan et al., 2009; Loftus & Palmer, 1974; Loftus & Pickrell, 1995). In one study (Chan & LaPaglia, 2013), participants were asked to watch an episode of a TV show. Afterwards, their memories were either reactivated or not before being presented with misinformation about the episode. In a subsequent recognition test, participants were impaired in their ability to recognize accurate information about the episode only if they were in the reactivation group. The studies in this dissertation show, however, that the interference from misinformation is more widespread than a third party attempting to introduce false information into one's memory. Misinformation can be produced by the participants themselves. In Section 2 Experiment 1, participants' exemplar recognition memory was impaired for objects that were included in the preceding intervening recall phase, suggesting that misinformation was spontaneously introduced during verbal recall, leading to its incorporation into memory. This then resulted in worse recognition memory performance for exemplar memory in the final recognition phase. While some may argue that the memory impairment on exemplar memory could be explained without misinformation, Section 4's results show clear evidence that misinformation was included in verbal recall, ultimately leading to both judges and participants being more likely to select the within-category foil over the target in a 2AFC task. Rather than simply being uninformative, the verbal descriptions were misleading.

All in all, our findings highlight that misinformation can affect memory without the experimenter needing to introduce the misinformation. With a simple testing procedure where participants are asked to recall what they had seen before, misinformation can be introduced by the participants themselves, subsequently impairing memory performance. This vulnerable nature of people's memories, where misinformation is spontaneously introduced by people during memory retrieval, had mainly been found in children (Brainerd & Reyna, 1998a; Reyna, 1995). To the best of our knowledge, our studies are some of the first few showing that spontaneous introduction of misinformation can also occur in adults.

5.1.3.2. Dual effects of recounting

Our studies also present the first evidence of a dual effect of recounting on memory. Research on how recounting affects memory has typically either reported a testing effect (i.e., benefit) from intervening retrieval (e.g., Roediger & Karpicke, 2006) or interference from recounting (e.g., Schooler & Engstler-Schooler, 1990; Chan et al., 2009). Our study shows that recounting can have multiple effects on visual memory, and that the relative strength of these effects changes across time. Section 2, for instance, shows that recounting leads to a net negative effect in the short term, while leading to a net positive effect in the long term with the inclusion of a delay after recounting. In other words, how intervening recounting affects subsequent memory depends on the relative strength of the opposing effects. These results present the clearest indication of recounting having multiple effects on memory. The data also confirm what has been observed in the testing effect literature, specifically supporting the idea that recounting benefits memory by slowing down forgetting (retrieval effort account; Carpenter & DeLosh, 2006). Because the benefit to memory is from slowed forgetting, it is no surprise that a net positive effect from recounting is only observed with a delay after intervening recall, once forgetting of the not-recalled material has occurred.

5.1.3.3. Visual memory processes

Besides shedding light on the dynamic nature of intervening retrieval's effects on visual memory, our experimental design also allows for the separate examination of how intervening retrieval affects category versus detail visual memory. In doing so, our data uncover specific processes underlying visual memory. Category information was never impaired by intervening recall (Sections 2 and 3), suggesting that category memory is less vulnerable to the detrimental effects of verbal re-coding and misinformation, an idea that aligns with past findings. Past studies found memory enhancement from intervening retrieval when memory was tested in a coarser manner and at the more verbalizable level (e.g., Carpenter & Pashler, 2007). The details of visual memory, however, may be more susceptible to impairments from verbalization, particularly in the short term (Section 2). The greater difficulty of remembering details compared to category information, especially over time (Brainerd, Reyna, & Ceci, 2008), suggests that detail visual memory may be more susceptible to inaccuracies, leading to impaired memory performance later on via reactivation and reconsolidation effects, for instance. The attempt to verbalize a more visual component of visual memory (i.e., the details) may also have interfered with memory through verbal overshadowing, affecting subsequent memory performance in the short term.

5.2. Usefulness of verbal descriptions for others

5.2.1. Findings

In Section 4, the utility of one's verbal descriptions of visual objects to others was examined. Participants either had to describe objects from perception (i.e., while looking at the object) or from memory so that someone else (i.e., the judges) could identify the object they saw based on their descriptions. When objects were described from perception, participants were able to provide fairly useful descriptions to judges, with judges identifying the target objects about 87% of the time. However, judges' accuracy dropped drastically to near chance levels when participants described the objects from memory. Analysis of the participants' own recognition accuracy revealed that the usefulness of their verbal descriptions heavily depended on whether the participants themselves had good memories for the objects. When participants' memories were good, leading to accurate responses during the recognition test, their prior descriptions also led judges to perform better for those objects. When they made recognition errors, however, their descriptions also misled judges to make errors. Furthermore, participants who had better recognition accuracy overall tended to provide the judges with better descriptions, leading to higher judges' accuracy. This correspondence between the participants' recognition accuracy and the usefulness of the verbal descriptions they provided suggests that verbal recall and visual identification may share common memory sources.

5.2.2. Significance of the findings

The findings grant insight into an important point regarding verbal reports, which is that some information is lost during one's translation of their visual representations into verbal reports. When describing from perception, one might predict that judges' accuracy would be close to or at 100%, since memory is not a limiting factor during description. However, judges performed at 87% accuracy, significantly below perfect performance. This finding suggests that participants may have limited descriptive ability overall. In addition, a more obvious reason for the difference between judges' and participants' accuracy is that participants did not know what the foils were going to be. As a result, they were not aware of which features of the object were diagnostic, and therefore, important to describe. In the real world, it is not uncommon to be unaware of what the foils are when describing what was seen to someone else. As such, the findings in this study show a possible limitation of one's ability to describe an object well for someone else in the real world. These limitations also played out when participants were describing the objects from memory. When participants' memories for the objects were good so that they correctly chose the target later on, their own recognition accuracy was still significantly higher than the judges' accuracy. These results, therefore, also suggest that participants have more information about the target in their memories than they are able to verbally describe.

The presence and role of misinformation are also highlighted in Section 4's findings. While it is possible that when participants' memories were poor, they simply did not provide enough information for judges to perform well, there is evidence that misinformation was indeed provided during recounting in some situations. When participants later chose a between-category foil, their prior verbal descriptions led judges to perform at chance, particularly when the verbal descriptions were provided 24 hours after the objects were seen (Section 4 Experiment 3). In this case, their verbal descriptions may have lacked useful information and been too sparse for judges to decide between the target and within-category foil. When participants later chose the within-category foil, however, their prior verbal descriptions for those objects led judges to perform significantly below chance. In other words, the verbal descriptions were not just unhelpful – they were misleading. These results suggest that misinformation was included in those descriptions, eventually leading judges to choose the within-category foil more often than the target, and possibly also affecting the participants' own subsequent recognition accuracy by interfering with memory. These results highlight a critical implication for the real world, particularly in criminal investigations. When participants were more likely to choose the within-category foil, they also misled judges to do the same. In other words, internal consistency between the participants' (or witness') and judges' accuracies does not imply that the choice made was accurate.

The results in Section 4 also have theoretical implications for memory more broadly. Past research has shown a poor correspondence between recall and recognition (Pozzulo et al., 2007), suggesting that recall and recognition may rely on different sources of memory. This argument is problematic for a U.S. Supreme Court guideline (Neil v. Biggers, 1972) for evaluating identification evidence, which states that a more accurate prior description should suggest that a subsequent identification made by a witness is also more reliable. In Section 4, where the verbal description task was a recall task (Experiments 2 and 3) and the subsequent identification task was a recognition task, a correspondence between the participants' recognition accuracy and the usefulness of their prior verbal descriptions was found. Participants who had better recognition accuracy overall also gave better descriptions to judges. In addition, for any given participant, if they later accurately identified an object, they also gave a better description for that object to judges. These findings lend support to the argument that verbal recall and visual identification may in fact rely on similar sources of memory. This also suggests that participants' subsequent identification accuracy is predictive of how useful their prior descriptions will be for others, supporting the guideline from the U.S. Supreme Court. One reason why our results may differ from past findings (Pozzulo et al., 2007) could be due to the type of stimuli used in our study (i.e., objects) versus their study (i.e., faces). Objects are more verbalizable than faces, increasing the likelihood that a more linguistic task such as recall and an arguably more visual task, recognition, would still tap into common memory sources. To gain deeper insight into the reason behind this discrepancy, future research should specifically examine how the correspondence between recall and recognition differs for different types of stimuli. Based on the current findings, however, it seems that in deciding how predictive memory performance in one task is for another memory task, people need to look carefully at the type of stimuli the person is asked to remember.

Overall, the data from Section 4 paint a grim picture of how well people can actually convey information about what they have seen to others. But how much can we generalize these findings? Are there situations where utility of verbal reports would be higher than that reported here? The usefulness of verbal reports may, first of all, be higher for categorical information. Categorical information of what was seen is commonly provided by people before the more detailed information. This, in addition to the more diagnostic nature of categorical information, likely increases the usefulness of verbal reports for identifying the right categories of objects that were seen by the describer.

Another situation in which verbal report utility may be higher than what was found in Section 4 is in the case of more familiar third parties and situations. In our study, participants were presented with individual, generic, objects and were told to describe them for someone else. They were not aware of who the third party was going to be. Judges, likewise, were not aware of who the participants were. Therefore, they had to base their decisions entirely on the participants' verbal descriptions and nothing else. In the real world, however, there are many cases when the person we are speaking to and providing the description for is familiar to us, such as a family member. The context in which an object is seen may also be familiar (e.g., a grocery store that one's family frequents). In these cases, we are likely able to tap into shared knowledge between ourselves and that family member, highlighting what we know to be diagnostic features of the object we are describing. In some cases, we may even be able to simply state a label or name of a product and the family member will know exactly which item to get while grocery shopping. This prediction is supported by past work on the phenomenon of common ground, which shows that the more shared knowledge people have, the fewer words they need to communicate with each other and direct the other person to a specific picture they are referring to in a matching task, for example (e.g., Krauss & Weinheimer, 1966; Isaacs & Clark, 1987). Hence, in these cases, it is likely that the utility of verbal descriptions, even those from memory, are much higher than is reported in Section 4.

Having said this, cases similar to what is depicted in Section 4 are not uncommon either. When providing directions to a stranger, or to a family member in a foreign place, or even in higher stakes situations such as after witnessing a crime, we are often called to describe what was seen from memory to another person who does not possess shared knowledge of the environment and the objects in it. In these cases, the utility of verbal descriptions may be lower than that reported in Section 4. In the absence of product labels, needing to describe things that are less verbalizable than common objects (e.g., faces), and not knowing what the "foils" would be, of which there may certainly be more than one, utility of one's verbal descriptions likely lowers even further than that observed in Section 4. This would then lead to at or below chance performance of "judges" in the real world. Our results, therefore, highlight the importance of relying on surveillance records rather than verbal reports whenever possible, since the utility of verbal reports can be low and may even be misleading. Even a mere 24 hours after encoding, verbal descriptions become almost useless to third parties (Section 4 Experiment 3).

The idea of surveillance may lead some to argue that with the prevalence of smartphones, people can take pictures and videos rather than verbally describe what they saw to someone else, effectively decreasing the need to be worried about the low utility that verbal reports have for third parties. With the increasing number of smartphones, however, people have increased access to social media, email, and applications that still involve describing what one has seen. When police are trying to locate a perpetrator, for instance, they often post a description of the person, based on what the eyewitness had told them. The public is then asked to use this description to locate the perpetrator. Easy access to the internet via smartphones means that the verbal description of the person is able to reach more people in a shorter amount of time than ever before. The speed at which information is transmitted to the wider public with the rise of smartphones is also astounding. While one might expect that a short delay between witnessing a crime and posting a description would result in an accurate description, our study shows that even after a 2-minute delay, verbal descriptions can drop drastically in utility and even be misleading. If we are able to increase our awareness of the limitations of verbal descriptions in cases such as these, we can work towards reducing the number of false convictions that result from faulty eyewitness testimony.

Might there be a way for us to be trained to become expert describers? Without the aid of visuals for the listener, sports commentators, for instance, have to describe the game they are watching over the radio. Their expert ability to describe the key moments of a game in real time raises the possibility that providing a useful description of what one saw, or is seeing, is possible. Before categorizing these commentators as expert describers, however, it is important to more closely examine the situations they are typically in. Several aspects of those situations make verbally describing a game easier than that presented in Section 4. Firstly, commentators are able to describe from perception (i.e., while seeing the game). Memory is rarely a limitation, unless the commentary occurs after a game. Even then, playback videos are usually available. As seen from Section 4, removing the limitation of memory can greatly increase the usefulness of one's verbal description for third parties. Secondly, the audience (i.e., the third parties) often have shared knowledge with the commentator about the layout of the field, the markings important to the game, and the roles that the players have on a team. These, in addition to having verbalizable labels such as jersey numbers, players' names, and field markings (e.g., "the 50 mark"), lead to an increased ease of describing a game for a radio listener. In these situations, it is uncommon to need to describe unmarked, generic objects or persons, since those are unlikely to be important to the game. If sports commentators are placed in a situation where they have to be eyewitnesses instead, it is unclear whether they will be able to provide verbal descriptions of the perpetrator and the objects present at a scene more effectively than the general public.

The strikingly low utility of verbal descriptions in the situation presented in Section 4, where there is brief exposure to multiple objects and immediate or delayed recall, begs the question of whether there are ways to improve third party accuracy from near chance to at least 87% (the accuracy of judges when verbal descriptions were provided from perception, Section 4 Experiment 1). Past work (Christie & Ellis, 1981) which found fairly good judges' accuracy from verbal descriptions points to a couple of possible ways to improve verbal report utility. Participants were first shown a target face for 60s before being asked to verbally describe the face. If they missed out on one of the main features of the face in their description (hair/eyes/nose/lips/chin), the experimenter prompted them to provide information for that feature. Judges subsequently used these participants' descriptions to select the target face from a lineup. Judges were able to identify about half of all the target faces. As such, it is possible that a longer encoding time for describers would increase the utility of the verbal descriptions for third parties. It should be noted, however, that the time one has to encode a visual stimulus is not always within their control. Furthermore, even with a prolonged viewing time of a visual stimulus, people have been found to make grave errors during identification ("Innocence Project", n.d.). As such, the reliability and usefulness of a verbal description should not be assumed just because encoding time was long. The effectiveness of extended encoding times is likely modulated by how describable and easily remembered the visual stimulus is. Another way in which the procedure in Christie and Ellis (1981) differed from ours is by prompting participants for specific feature descriptions. Doing this during the collection of verbal reports may therefore increase the utility of verbal descriptions for others. Having said this, precautions should also be taken with this intervention. One should be careful about the possibility of describers churning out misinformation in an effort to provide descriptions for all the key features of what was seen. As shown in Section 4, misinformation can not only mislead judges to make errors, it may lead to memory interference for the describers themselves, resulting in worse subsequent recognition accuracy.

The continued examination of the utility of verbal reports, beyond what is reported in this dissertation, is important to expand the field's understanding of the nature of memory, but also develop ways in which utility can be increased for third parties, especially in higher stakes situations like during eyewitness testimony.

5.3. Delay between encoding and intervening recall

In Sections 2 and 3, the testing benefit from intervening recall clearly emerged when there was a 24-hour delay between intervening recall and final recognition testing. These results align with past findings showing that the testing effect increases across a delay because intervening retrieval slows forgetting of the material, supporting the retrieval effort account (Carpenter & DeLosh, 2006). Another prediction of the retrieval effort account is that the benefit from testing increases with retrieval effort on the intervening test. One way in which studies involving verbal materials (Karpicke & Roediger, 2007; Whitten & Bjork, 1977; Glover, 1989; Cuddy & Jacoby, 1982) have demonstrated this effect is by inserting a delay between initial encoding and intervening retrieval. Delaying the first retrieval practice makes the solution less accessible, and hence increases retrieval effort on that first retrieval practice, enhancing performance in subsequent tests (Cuddy & Jacoby, 1982). Thus, by inserting a delay between encoding and intervening recall, the retrieval effort account predicts that the benefit derived from intervening recall will increase.

Having said this, retrieval benefits have typically been shown to depend on there being successful, fruitful retrieval of the material during the intervening test (Liu & Reder, 2016; Kornell et al., 2011; Rowland, 2014). As such, it is possible that introducing a delay between encoding and intervening retrieval will lead to enough memory decay so that retrieval during the intervening test will not reach a sufficient level of fruition. If so, subsequent memory performance could either show no benefits from intervening recall because people are not able to retrieve any useful information during intervening recall, or memory interference because misinformation is introduced during intervening recall.

Data reported in this section are from an unpublished study where a 24-hour delay was inserted between Phase 1 (encoding) and Phase 2 (intervening recall) to examine these effects.

5.3.1. Method

5.3.1.1. Participants. Twenty-four participants completed this study. There were 19 females and 5 males, with a mean age of 19.3 years.

5.3.1.2. Design and procedure. This experiment was identical to that in Section 2 Experiment 4A, except that the 24-hour delay occurred between the first (encoding) and second (intervening recall) phases, rather than between the second and third (final memory test) phases.

5.3.2. Results

As before, both the effect of intervening recall on category memory and exemplar memory were analyzed. The modulating effect of delay was also analyzed, with delay interval as a between-subject factor in subsequent analyses.

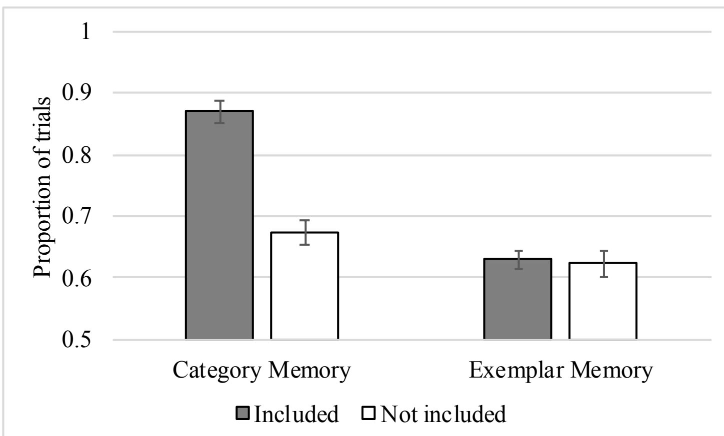

Figure S5.1. Recognition category and exemplar memory results for objects included and not included in the intervening recall phase when there was a 24-hour memory delay between Phases 1 and 2. Error bars showed ±1S.E. of the mean.

5.3.2.1. Immediate (Section 2, Experiment 1) vs. 24-hour delay between Phases 1 and 2 (current experiment)

Category memory was enhanced for objects that were included in the intervening recall phase (Figure S5.1.). This enhancement was, however, stronger when there was a 24-hour delay between Phase 1 and 2, compared to when all phases occurred on the same day. An ANOVA with intervening recall as a within-subject factor and memory delay as a between-subject factor showed a significant main effect of intervening recall, $F(1, 59) = 305.20$, $p < .001$, $\eta_p^2 = .84$. Intervening recall increased category memory by 19.63% with a delay, compared to 12.37% without a delay.

The increased memory enhancement from the delay, however, was not observed with exemplar memory (Figure S5.1.). An ANOVA with intervening recall as a within-subject factor and memory delay as a between-subject factor did not show any significant main effect or interaction, $F(1, 59) = 2.82$, $p = .098$. However, a comparison of exemplar memory performance for objects that were included ($M = 63.02\%$, $S.E. = 1.4$) and not included in intervening recall ($M = 62.31\%$, $S.E. = 2.0$) showed that performance was similar between the two conditions, $t(23) = .035$, $p = .73$, unlike that observed when all three phases occurred on the same day (Section 2, Experiment 1).

5.3.2.2. 24-hour delay between Phases 1 and 2 (current experiment) vs. between Phases 2 and 3 (Section 2, Experiment 4)

The effects of intervening recall on category memory and exemplar memory, when the 24-hour delay was inserted between Phase 1 (encoding) and Phase 2 (intervening recall) or between Phase 2 and Phase 3 (final memory test), were compared. As observed when the 24-hour delay was inserted between Phase 2 (intervening recall) and Phase 3 (final memory test) in Section 2 Experiment 4, inserting a 24-hour delay between Phases 1 and 2 led to an enhancement of category memory for objects included in intervening recall compared to those that were not. An ANOVA with intervening recall (included and not included) as a within-subject factor and Experiment as a between-subject factor showed a significant main effect of intervening recall, $F(1, 46) = 395.63$, $p < .001$, $\eta_p^2 = .90$. There was no interaction between intervening recall and Experiment, $F(1, 46) = 2.47$, $p = .12$.

The effects of intervening recall on exemplar memory, however, differed depending on when the 24-hour delay was introduced. Whereas a 24-hour delay between Phases 2 and 3 led to memory enhancement from intervening recall, a delay between Phases 1 and 2 in this experiment led to similar memory performance during the final memory test for objects that were included and not included in intervening recall. An ANOVA on intervening recall (included and not included) and Experiment revealed a significant main effect of intervening recall, $F(1, 46) = 6.86$, $p = .012$, $\eta_p^2 = .13$, and a significant interaction, $F(1, 46) = 4.26$, $p = .045$, $\eta_p^2 = .085$. Hence, intervening recall enhanced exemplar memory by about 6% when the 24-hour delay was between Phases 2 and 3, but did not lead to any facilitation when the 24-hour delay was between Phases 1 and 2.

5.3.3. Discussion

Inserting a 24-hour delay between the encoding and intervening recall phases produced a pattern of results largely similar to that observed when the delay was between intervening recall and final memory test, but only for category memory. Intervening recall increased category memory when there was a delay introduced between Phases 1 and 2, compared to when all three phases occurred on the same day. These results were predicted by the retrieval effort account, suggesting that intervening recall leads to a testing benefit because testing increases elaborative retrieval processing and thus, accessibility to the material. As in Section 2, however, this increase in category memory from intervening recall may also have been due to the nature of the intervening recall test. The cue used in this experiment was a category cue, which re-exposed participants to the objects' category name. As such, it is possible that the category memory facilitation was due to category information re-exposure rather than recall itself, or both.

Unlike for category memory, intervening recall did not significantly affect exemplar memory. Exemplar memory was not significantly different for objects that were included in intervening recall compared to those that were not included. This is in contrast to the condition where the 24-hour delay was inserted between Phases 2 and 3 (Section 2 Experiment 4), which found an exemplar memory facilitation from intervening recall. The lack of exemplar memory facilitation in the current experiment is surprising, given the memory facilitation prediction from the verbal literature (Karpicke & Roediger, 2007; Whitten & Bjork, 1977; Glover, 1989; Cuddy & Jacoby, 1982). One reason for this discrepancy could be the fact that memory decays over time. While the retrieval effort account predicts that a delay before intervening testing would lead to greater retrieval effort, and therefore, even greater memory enhancement from intervening testing, this facilitation has been found to depend on whether retrieval is successful (Liu & Reder, 2016; Kornell et al., 2011; Rowland, 2014). In the current experiment, it is possible that memory for the object details decayed enough that during the intervening recall 24 hours later, participants were not able to recall enough information about the objects to strengthen the memory trace. This explanation aligns with what was found for category memory. The category memory enhancement from intervening recall was larger in the current study compared to when all three phases occurred on the same day. This is likely due in part to the re-exposure to the category name, as well as the ease of remembering gist information compared to detailed information (Brainerd et al., 2008). When category information of the objects was retrieved, this strengthened category memory by a greater magnitude with a delay than without one.

Interestingly, an exemplar memory detriment was not observed in this current study, as was found in the same-day condition (Section 2 Experiment 1). A detrimental effect of intervening recall with a delay is partly predicted by the verbal overshadowing literature (Alogna et al., 2014, RRR2). Participants in their study watched a video of a robbery, completed a filler task for 20 minutes (i.e., had a delay), and then either wrote out a verbal description of the perpetrator or did another unrelated filler task. In a final recognition task immediately after the verbal description, participants had to pick out the perpetrator from a lineup. As in our study, the delay was inserted between encoding and verbal recall. Participants in the verbal overshadowing study were 16% less likely to select the target if they had completed the intervening verbal recall task than if they had not. Our study, however, did not show such a detrimental effect from intervening recall. Instead, it seems that recalling objects was as good as not having recalled them at all when there was a 24-hour delay between encoding and intervening recall test. The verbal overshadowing account suggests that verbalization suppresses harder to verbalize features. This then leads to worse performance in a subsequent recognition task, which depends in large part on remembering the less verbalizable, more visual, features of what was seen. It is possible, therefore, that after a 24-hour delay, memory was too poor for there to be "sufficient" visual memory to suppress. This and not being able to retrieve sufficient information for a fruitful retrieval are likely the reasons that intervening recall had little effect on exemplar memory compared to the no-recall condition. All in all, it appears that a good balance between increasing retrieval effort, by increasing the difficulty of the intervening test, and having successful retrieval are necessary for testing benefits to emerge from intervening testing. Future research should modify the length of the delay to be shorter than 24 hours to see whether that would increase fruitful retrieval while still increasing retrieval difficulty enough to observe a testing benefit from intervening recall with the type of stimuli used in our study.

5.4. Conclusion

The act of recounting one's memory of what was seen is prevalent in daily life, making the investigation of how it can affect our own memories, and others' actions, essential. This dissertation reveals the dynamic nature of memory, suggesting that the way in which recounting affects our memories varies and depends on a number of specific factors. The effect of recounting depends on the type of memory (category or exemplar, item or source) one is considering, whether a delay is involved, when the delay occurs, and the type of stimuli one is recounting, to name a few. Mechanistically, the success and amount of retrieval effort exerted during recounting can significantly determine the extent to which recounting results in a memory facilitation or impairment as well. The discrepancy between impressive visual memory and unreliable visual memory in past research can therefore be accounted for, at least in part, by the act of recounting one's memory.

This dissertation also makes several suggestions with regards to the ways in which memory can be facilitated or impaired through recounting. In terms of memory enhancement, the retrieval effort account presents a strong case for the mechanism underlying the facilitation. When it comes to memory impairment from recounting, verbal re-coding and interference from misinformation are two ways in which recounting seems to impair visual memory. This is not to say, however, that other explanations for the memory facilitation and impairment are invalid. In future efforts to map out the nature of memory and determine the way it works, researchers should consider all theories of memory presented in the field, and bear in mind that memory is complex and dynamic, and may have different mechanisms at play depending on the situation. A continued effort to refine our understanding of which mechanisms underlie memory in different situations would be invaluable to comprehending the complexity of memory.

This dissertation also examined the common act of recounting our memories in social communication. In situations where familiarity and shared knowledge are absent, coupled with brief exposure to what was seen, our verbal descriptions are, unfortunately, of limited utility to others. This limitation of verbal reports is important to keep in mind when deciding how much weight to place on verbal reports of what was seen, especially in higher stakes situations. Further research into the mechanisms underlying the translation of one's visual memory to others can lead to a greater understanding of not only how this translation can be made more effective, but also help determine when a verbal description is more reliable. By examining the correspondence between one's verbal report and their subsequent ability to remember what was seen again, we can also gain greater insight into the nature of memory itself.

What might account for the discrepancy between the effects of recounting on the describers (i.e., the participants) versus the listeners (i.e., the judges)? Why does recounting lead to benefits in one's memory in the long term, but have limited usefulness for someone else? One possible reason is that recounting activates an experience in the describers such that it serves to strengthen those memories during recounting. In third parties, however, these corresponding experiences do not exist. Therefore, the recounted memories do not activate an internal representation in judges like they might do in participants. The judges' need to rely completely on the participants' descriptions in their selections, making them more vulnerable to the detrimental effects that recounting can bring. Any misinformation included in the descriptions would likely affect judges more than participants due to the absence of any other memory representations, and therefore, cues, in the judges' minds. Further research is needed to confirm whether these ideas hold true across all situations.

As future researchers embark on these questions of how memory works and build on what has already been found, it is crucial to keep in mind the importance of the reproducibility of scientific findings (Open Science Collaboration, 2015). Without replicability, scientific findings should not be treated as the be all and end all, but rather as steps that bring us closer to the truth. The quest to understand the nature of memory is a daunting, but exciting, one. The findings in this dissertation grant some clarity about the dynamic nature of our memories and serve as another stone on which future research can build.

Bibliography

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861-876. doi:10.1002/acp.1391

Akirav, I., & Maroun, M. (2006). Ventromedial prefrontal cortex is obligatory for consolidation and reconsolidation of object recognition memory. *Cerebral Cortex, 16*, 1759-1765. doi:10.1093/cercor/bhj114

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, S., Birch, S., Birt, A. R., ...Zwaan, R. A. (2014). Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science, 9*, 556-578. doi:10.1177/1745691614545653

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review, 7*, 522-530. doi:10.3758/BF03214366

Anderson, A. K., Wais, P. E., & Gabrieli, J. D. E. (2006). Emotion enhances remembrance of neutral events past. *Proceedings of the National Academy of Sciences, 103*, 1599-1604. doi:10.1073 pnas.0506308103

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*, 66-80. doi:10.1162/jocn.1995.7.1.66

Bartlett, F. C. (1932). *Remembering: An experimental and social study.* Cambridge: Cambridge University.

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

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*, 173-189. doi:10.1016/0010-0285(74)90009-7

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*, 14325-14329. doi:10.1073/pnas.0803390105

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436. doi:10.1163/156856897X00357

Brainerd, C. J., & Reyna, V. F. (1998a). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71*, 81-129. doi:10.1006/jecp.1998.2464

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin, 134*, 343-382. doi:10.1037/0033-2909.134.3.343

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik

(Eds.), *Levels of processing in human memory* (pp. 331-354). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory, 18*, 385-393. doi:10.1080/09658211003702163

Brown, C., & Lloyd-Jones, T. J. (2005). Verbal facilitation of face recognition. *Memory & Cognition, 33*, 1442-1456. doi:10.3758/BF03193377

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 7*, 207-218. doi:10.1037//1076-898X.7.3.207

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619-636. doi:10.1002/acp.1101

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. doi:10.3758/BF03193405

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review, 14*, 474-478. doi:10.3758/BF03194092

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1563-1569. doi:10.1037/a0017021

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36*, 438-448. doi:10.3758/MC.36.2.438

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633-642. doi:10.3758/BF03202713

Ceci, S. J., Loftus, E. F., Leichtman, M. D., & Bruck, M. (1994). The possible role of source misattributions in the creation of false beliefs among preschoolers. *The International Journal of Clinical and Experimental Hypnosis, 42*, 304-320. doi:10.1080/00207149408409361

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. doi:10.1073/pnas.1218472110

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility. *Psychological Science, 20*, 66-73. doi:10.1111/j.1467-9280.2008.02245.x

Christie, D. F. M., & Ellis, H. D. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology, 66*, 358-363. doi:10.1037/0021-9010.66.3.358

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1403-1414. doi:10.1037/0278-7393.25.6.1403

Craik, F. I. M., & Jennings, J. M. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 51-110). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior, 21*, 451-467. doi:10.1016/S0022-5371(82)90727-7

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 2157-2162. doi:10.1073/pnas.1218472110

Davies, G. (1986). Capturing likeness in eyewitness composites: The police artist and his rivals. *Medicine, Science and the Law, 26*, 283-290. doi:10.1177/002580248602600407

Diekelmann, S., Büchel, C., Born, J., & Rasch, B. (2011). Labile or stable: Opposing consequences for memory when reactivated during waking and sleep. *Nature Neuroscience, 14*, 381-286. doi:10.1038/nn.2744

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology, 55*, 51-86. doi:10.1146/annurev.psych.55.090902.142050

Duvarci, S., & Nader, K. (2004). Characterization of fear memory reconsolidation. *The Journal of Neuroscience, 24*, 9269-9275. doi:10.1523/JNEUROSCI.2971-04.2004

Eisenberg, M., Kobilo, T., Berman, D. E., & Dudai, Y. (2003). Stability of retrieved memory: Inverse correlation with trace dominance. *Science, 301*, 1102-1104. doi:10.1126/science.1086881

Elsey, J. W. B., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin, 144*, 797-848. doi:10.1037/bul0000152

Fan, J., Snodgrass, J. G., & Bilder, R. M. (2003). Functional magnetic resonance imaging of source versus item memory. *Brain Imaging, 14*, 2275-2281. doi:10.1097/01.wnr.0000090582.35425.e1

Finger, K., & Pezdek, K. (1999). The effect of the cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology, 84*, 340-348. doi:10.1037/0021-9010.84.3.340

Finn, B., & Roediger, H. L. (2011). Enhancing retention through reconsolidation: Negative emotional arousal following retrieval enhances later recall. *Psychological Science, 22*, 781-786. doi:10.1177/0956797611407932

Finn, B., Roediger, H. L., & Rosenzweig, E. (2012). Reconsolidation from negative emotional pictures: Is successful retrieval required? *Memory & Cognition, 40*, 1031-1045. doi:10.3758/s13421-012-0203-7

Gates, A. I. (1917). Recitation as a factor in memorizing. In R. S. Woodworth (Ed.), *Archives of psychology* (pp. 1-104). New York: The Science Press.

Glisky, E. L., Polster, M. R., Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9*, 229-235. doi:10.1037/0894-4105.9.2.229

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392-399. doi:10.1037/0022-0663.81.3.392

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*, 330-337. doi:10.1016/S1364-6613(00)01519-9

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1997). Task constraints in visual working memory. *Vision Research, 38*, 125-137. Retrieved from https://www.journals.elsevier.com/vision-research

Hess, T. M., & Slaughter, S. J. (1990). Semantic knowledge influences on memory for scene information in young and older adults. *Developmental Psychology, 26*, 855-865. doi:10.1037/0012-1649.26.5.855

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 113-136. doi:10.1037//0096-1523.28.1.113

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 8*, 761-768. doi:10.3758/BF03196215

Innocence Project. (n.d.). Ronald Cotton. Retrieved from http://www.innocenceproject.org/cases-false-imprisonment/ronald-cotton

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116*, 26-37. doi:10.1037/0096-3445.116.1.26

Jackson, M. C., & Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology: Human Perception & Performance, 34*, 556-568. doi:10.1037/0096-1523.34.3.556

Jackson, M. C., & Raymond, J. E. (2006). The role of attention and familiarity in face identification. *Perception & Psychophysics, 68*, 543-557. doi:10.3758/BF03208757

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*, 528-558. doi:10.1080/09541440601056620

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*, 151-162. doi:10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966-968. doi:10.1126/science.1152408

Kindt, M., Soeter, M., & Vervliet, B. (2009). Beyond extinction: Erasing human fear responses and preventing the return of fear. *Nature Neuroscience, 12*, 256-258. doi:10.1038/nn.2271

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*, 1551-1556. doi:10.1177/0956797610385359

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65*, 85-97. doi:10.1016/j.jml.2011.04.002

Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: Testing the semantic categorization account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 499-510. doi:10.1037/0278-7393.29.4.499

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343-346. doi:10.1037/h0023705

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4,* 501-506. doi:10.3758/BF03214339

Lindholm, T., & Christianson, S. Å. (1998). Intergroup biases and eyewitness testimony. *The Journal of Social Psychology, 138*, 710-723. doi:10.1080/00224549809603256

Litvin, O. O., & Anokhin, K. V. (2000). Mechanisms of memory reorganization during retrieval of acquired behavioral experience in chicks: The effects of protein synthesis inhibition in the brain. *Neuroscience and Behavioral Physiology, 30*, 671-678. doi:10.1023/A:1026698700139

Liu, K., & Jiang, Y. (2005). Visual working memory for briefly presented scenes. *Journal of Vision, 5*, 650-658. doi:10.1167/5.7.5

Liu, X. L., & Reder, L. M. (2016). fMRI exploration of pedagogical benefits of repeated testing: When more is not always better. *Brain and Behavior, 6*, e00476. doi:10.1002/brb3.476

Liu, X. L., Liang, P., Li, K., & Reder, L. M. (2014). Uncovering the neural mechanisms underlying learning from tests. *PLOS ONE, 9*, e92025. doi:10.1371/journal.pone.0092025

Liu, X. L., Tan, D. H., & Reder, L. M. (2018). The two processes underlying the testing effect - Evidence from event-related potentials (ERPs). *Neuropsychologia, 112*, 77-85. doi:10.1016/j.neuropsychologia.2018.02.022

Loftus, E. F., & Burns, T. E. (1982). Mental shock can produce retrograde amnesia. *Memory & Cognition, 10*, 318-323. doi:10.3758/BF03202423

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585-589. doi:10.1016/S0022-5371(74)80011-3

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25*, 720-725. doi:10.3928/0048-5713-19951201-07

Loftus, E. F. (1979). The malleability of human memory: Information introduced after we view an incident can transform memory. *American Scientist, 67*, 312-320. Retrieved from http://www.americanscientist.org/

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

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

Maxcey, A. M., & Woodman, G. F. (2014). Forgetting induced by recognition of visual images. *Visual Cognition, 22*, 789-808. doi:10.1080/13506285.2014.917134

Meissner, C. A. (2002). Applied aspects of the instructional bias effect in verbal overshadowing. *Applied Cognitive Psychology, 16*, 911-928. doi:10.1002/acp.918

Meissner, C. A., Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15*, 603-616. doi:10.1002/acp.728

Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology, 20*, 414-455. doi:10.1080/09541440701728581

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

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

Neil v. Biggers, 409 U.S. 188 (1972)

Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott false memory paradigm: The role of identifiability. *Journal* *of Experimental Psychology: Learning, Memory, and Cognition, 29*, 35-41. doi:10.1037/0278-7393.29.1.35

Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology, 11*, 287-307. doi:10.1016/0010-0285(79)90013-6

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*, 939-1031. doi:10.1017/S0140525X01000115

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716. doi:10.1126/science.aac4716

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 406-417. doi:10.1037/0278-7393.3.4.406

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442. doi:10.1163/156856897X00366

Pigott, M., & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology, 70*, 547-555. doi:10.1037/0021-9010.70.3.547

Pozzulo, J. D., Lemieux, J. M. T., Wells, E., & McCuaig, H. J. (2007). The influence of eyewitness identification decisions and age of witness on jurors' verdicts and perceptions of reliability. *Psychology, Crime & Law, 12*, 641-652. doi:10.1080/10683160500415539

Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Psychological Science, 5*, 649-663. doi:10.1177/1745691610388763

Ranganath, C., DeGutis, J., & D'Esposito, M. (2004). Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Cognitive Brain Research, 20*, 37-45. doi:10.1016/j.cogbrainres.2003.11.017

Rensink, R. A., O'Regan, K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368-373. doi:10.1111/j.1467-9280.1997.tb00427.x

Reyna, V. F. (1995). Interference effects in memory and reasoning: A fuzzy-trace theory analysis. In F. N. Dempster & C. J. Brainerd (Eds.), *New perspectives on interference and inhibition in cognition* (pp. 29-61). New York: Academic Press.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning – Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. doi:10.1111/j.1467-9280.2006.01693.x

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

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192-233. doi:10.1037/0096-3445.104.3.192

Rossato, J. I., Bevilaqua, L. R. M., Myskiw, J. C., Medina, J. H., Izquierdo, I., & Cammarota, M. (2007). On the role of hippocampal protein synthesis in the consolidation and reconsolidation of object recognition memory. *Learning & Memory, 14*, 36-46. doi:10.1101/lm.422607

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

Sapkota, R. P., Pardhan, S., & van der Linde, I. (2013). Manual tapping enhances visual short-term memory performance where visual and motor coordinates correspond. *British Journal of Psychology, 104*, 249-264. doi:10.1111/j.2044-8295.2012.02115.x

Sapkota, R. P., van der Linde, I., & Pardhan, S. (2015). How does aging affect the types of error made in a visual short-term memory 'object-recall' task? *Frontiers in Aging Neuroscience, 6*, 1-9. doi:10.3389/fnagi.2014.00346

Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., LeDoux, J. E., Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature, 463*, 49-52. doi:10.1038/nature08637

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36-71. doi:10.1016/0010-0285(90)90003-M

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memories. In D. Medin (Ed.), *Handbook of Learning and Motivation* (Vol. 37). Orlando, FL: Academic Press.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*, 776-778. doi:10.1126/science.1207745

Spencer, W. D., & Raz, N. (1994). Memory for facts, source, and context: Can frontal lobe dysfunction explain age-related differences? *Psychology and Aging, 9*, 149-159. doi:10.1037/0882-7974.9.1.149

St. Jacques, P. L., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences, 110*, 19671-19678. doi:10.1073/pnas.1319630110

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*, 73-74. doi:10.3758/BF03337426

Staresina, B. P., & Davachi, L. (2009). Mind the gap: Binding experiences across space and time in the human hippocampus. *Neuron, 63*, 267-276. doi:10.1016/j.neuron.2009.06.024

Summers, M. J., Crowe, S. F., & Ng, K. T. (1997). Administration of DL-2-amino-5-phosphonovaleric acid (AP5) induces transient inhibition of reminder-activated memory retrieval in day-old chicks. *Cognitive Brain Research, 5*, 311-321. doi:10.1016/S0926-6410(97)00007-4

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin & Review, 23*, 831-841. doi:10.3758/s13423-015-0937-x

Suzuki, A., Josselyn, S. A., Frankland, P. W., Masushige, S., Silva, A. J., & Kida, S. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *The Journal of Neuroscience, 24*, 4787-4795. doi:10.1523/JNEUROSCI.5491-03.2004

Tan, D. H., & Jiang, Y. V. (2018). How visual memory changes with intervening recall. *Quarterly Journal of Experimental Psychology*, 1-11. doi:10.1177/1747021818808524

Tan, D. H. & Jiang, Y. V. (2019). *Tell me what you saw: The usefulness of verbal descriptions for others.* Manuscript under review.

Tan, D. H., & Jiang, Y. V. (2019). *The generalizability of the testing effect across memory cues and attributes.* Manuscript in preparation.

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the "testing effect": A review. *Trends in Neuroscience and Education, 5*, 52-66. doi:10.1016/j.tine.2016.05.001

van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage, 78*, 94-102. doi:10.1016/j.neuroimage.2013.03.071

Van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21*, 661-685. doi:10.1023/A:1024812831576

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

Wells, G. L. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology, 70*, 619-626. doi:10.1037/0021-9010.70.4.619

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

Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Behavior, 16*, 465-478. doi:10.1016/S0022-5371(77)80040-6

Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: An fMRI study of the testing effect. *Neuropsychologia, 51*, 2360-2370. doi:10.1016/j.neuropsychologia.2013.04.004

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception, 16*, 747-759. doi:10.1068/p160747n

Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron, 44*, 889-898. doi:10.1016/j.neuron.2004.11.018

Yu, C. J., & Geiselman, R. E. (1993). Effects of constructing identi-kit composites on photospread identification performance. *Criminal Justice and Behavior, 20*, 280-292. doi:10.1177/0093854893020003005