2010-01-01

The Role of Generation and Monitoring Processes in Governing the Paradoxical Effects of Retrieval on Memory For Faces

Kyle Joseph Susa
University of Texas at El Paso, kjsusa@miners.utep.edu

THE ROLE OF GENERATION AND MONITORING PROCESSES IN GOVERNING THE PARADOXICAL EFFECTS OF RETRIEVAL ON MEMORY FOR FACES

KYLE JOSEPH SUSA

Department of Psychology

APPROVED:

____
Christian Meissner, Ph.D., Chair

____
Ana Schwartz, Ph.D.

____
Harmon Hosch, Ph.D.

____
Theodore Curry, Ph.D.

____
Matthew Scullin, Ph.D.

____
Patricia D. Witherspoon, Ph.D.
Dean of the Graduate School

To “Grandma Rita,” for your eternal inspiration

THE ROLE OF GENERATION AND MONITORING PROCESSES IN GOVERNING THE PARADOXICAL EFFECTS OF RETRIEVAL ON MEMORY FOR FACES

By

KYLE JOSEPH SUSA, B.S., M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

THE UNIVERSITY OF TEXAS AT EL PASO

May 2010

ACKNOWLEDGEMENTS

I would like to express my gratitude to several people who have supported me in completing my degree. First and foremost, I would like to thank my family for being a continual source of support. I could not have made it without their encouragement to keep moving forward. My sincere appreciation also goes to my advisor, Dr. Christian Meissner, and to many faculty members in the Psychology Department at UTEP who have sacrificed on my behalf. Drs. Roy Malpass, Osvaldo Morera, Harmon Hosch, Larry Cohn, and Ted Cooper are close colleagues and professors who have always had my best interests in mind. Finally, I want to thank Marcela Diaz for her continued support and patience. She has endured the dissertation process with me, and I share the rewards of accomplishment with her.

ABSTRACT

Verbal descriptions of faces can at times impair and at other times facilitate subsequent face identification accuracy. Three experiments were conducted from a retrieval-based theoretical perspective to determine the underlying cognitive processes that can account for these paradoxical findings. Results demonstrated that the verbal description-identification relationship is analogous to other domains of memory where an initial retrieval of memory can have both positive and negative effects on subsequent retrieval attempts. Results of Experiment 1 showed that verbal facilitation is a product of self-generated descriptions that enrich the semantic encoding of the original memory trace. Experiments 2 and 3 demonstrated that descriptions impair identification accuracy when participants are forced to confabulate details. The negative effects of confabulation can be alleviated through retrieval monitoring and source memory training during the initial and subsequent memory retrieval attempts. Theoretical and practical implications are discussed.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Eyewitnesses to crime play a vital role in our legal system in providing first-hand information about the appearance of suspected perpetrators. In fact, eyewitnesses are often relied upon for recalling information about a perpetrator at multiple stages during the justice process. While this information is often very useful in assisting law enforcement, research suggests that many factors can *misinform* an eyewitness' memory beyond their conscious recollection (Gerrie, Garry, & Loftus, 2005). Unfortunately, research demonstrates this theme is particularly pronounced in memory for an accused perpetrators' face when face identification procedures (i.e. photo-lineups) follow a verbal description (see Meissner, Sporer, & Schooler, 2006, for a review).

Schooler and Engstler-Schooler (1990) were the first to demonstrate the adverse effects of verbal descriptions on subsequent face identification accuracy and labeled the phenomenon the *verbal overshadowing effect*. The investigators presented participants with a short video of a staged bank robbery, after which participants were either asked to provide a verbal description of the perpetrator (description condition) or asked to complete an unrelated filler task (control condition). A recognition test was then assigned, whereby participants had to identify the perpetrator out of a target-present lineup. Participants who were part of the description condition were significantly less accurate on this task than control participants (38% vs. 64%).

Despite research on the negative effects of descriptions on face memory, recent research has demonstrated that verbal descriptions can also positively enhance face recognition. This positive effect is known as the *verbal facilitation effect*. In a recent meta-analytic review of the verbal description-identification relationship in face memory, Meissner, Sporer, and Susa (2008) identified three factors that differentiate standard verbal overshadowing paradigms from those of verbal facilitation. These factors include the number of faces that are encoded, the amount of time given to generate a verbal description, and the delay time between description and identification tasks. Even with an understanding of these paradigm differences, the precise cognitive mechanisms underlying overshadowing vs. facilitation effects have yet to be established. The current studies seek to address this issue and distill, from a basic memory perspective, why producing a verbal description of a face sometimes overshadows and at other times facilitates subsequent face memory. Theoretical perspectives in the basic literature that speak to this issue can be narrowed to two governing principles, namely: retrieval-based accounts (Meissner, Brigham, & Kelley, 2001) and transfer appropriate processing (Schooler, 2002; Tulving & Thompson, 1973). The current experiments focus on an assessment of a retrieval-based perspective, including the effects of generation (or repeated testing), output (response) criterion and the influence of retrieval monitoring (on initial retrieval vs. repeated-retrieval attempts) on subsequent memory. To understand the influence of verbalization on face memory better a discussion of the literature is examined below.

Verbal Overshadowing

Since the original verbal overshadowing study by Schooler and Engstler-Schooler (1990) numerous experiments have replicated the effect on face identification (e.g., Dodson, Johnson, & Schooler, 1997; Fallshore & Schooler, 1995). Other studies have even extended the verbal overshadowing effect to other domains such as wine tasting (Melcher & Schooler, 1996), decision-making (Wilson & Schooler, 1991), and visual imagery (Brandimonte, Hitch, & Bishop, 1992). In a meta-analysis of verbal overshadowing studies Meissner and Brigham (2001) found a significant verbal overshadowing effect (*Zr* = .12), concluding that participants who give a verbal description of a target face are 1.27 times more likely than control participants to *misidentify* the face in a photo-lineup identification test. While these results demonstrate the verbal overshadowing effect is a reliable phenomenon, the underlying mechanisms and cognitive processes that drive the effect remain under dispute.

Two major theories of verbal overshadowing, a retrieval-based interference account, and a transfer (in)appropriate processing account, have been predominantly supported within the literature. Each of the two major theories of verbal overshadowing offers empirical support for its viewpoint, and each theory caries merit. However, a single theory has not been universally accepted. One of the major theoretical explanations is retrieval-based interference, proposed by Meissner et al. (2001). Retrieval-based interference theory focuses on the contents of verbal descriptions. It claims that inaccurate verbal descriptions of a face can lead to self-generated misinformation, which in turn alters or interferes with the person's original memory. Subsequently, when witnesses are asked to identify or recognize the face they just described they will likely identify the face consistent with their misinformed memory rather than their original memory. The interfering effects of self-generated misinformation should not be present when the original stimulus can be easily verbalized or when the contents of the description are qualitatively different than the stimuli that the person is being asked to identify.

Numerous studies have been conducted that support retrieval-based interference (i.e. Finger & Pezdek, 1999; Meissner, Brigham, & Kelley, 2001, Meissner, et al., 2008). For instance, Meissner et al. (2001) found that instructional manipulations to the description tasks can alter the output criterion established by participants and thereby influence the percentage of errors they generate in their descriptions and subsequently their accuracy rates in identification performance. Specifically, Meissner et al. manipulated the quality of the descriptions given by participants and subsequently measured their accuracy on an identification test (see also Finger

& Pezdek, 1999). In their “forced” recall condition participants were forced to generate 25 details about a target face. Results indicated that these participants were significantly more likely to generate inaccurate details than participants “warned” to provide only details that they could confidently recall. Meissner et al. (2001) subsequently found that, as a function of having more incorrect details on the verbal description task, “forced” conditioned participants performed significantly worse on the identification task, both immediately and after a 30 min. delay, than participants in all other conditions. However, when “warned,” participants generated accurate descriptions and actually demonstrated a verbal facilitation effect.

The retrieval-based interference account is further supported by research demonstrating that verbal overshadowing is especially likely when perceptual expertise exceeds verbal ability (i.e. Melcher & Schooler, 1996; Ryan & Schooler, 1998). Accordingly, people who have a high level of perceptual expertise are able to perceive and process numerous featural aspects of a stimulus, however, they are not able to put these perceptions into words. Subsequent verbalization of the perception is inaccurate and unrepresentative of the original experience (Chin & Schooler, 2008). In a study examining the effect of varying levels of perceptual and verbal expertise, Melcher and Schooler (1996) asked non-wine drinkers, untrained wine drinkers, and trained wine experts to taste a wine, and then either verbalize it or perform a control task before being asked to identify the wine from a series of distractors. In support of a retrieval-based account they found that verbalization only impaired untrained wine drinkers, and did not affect trained wine drinkers or non-wine drinkers. The authors concluded that when perceptual and verbal abilities are approximately equivalent, as they would be in trained wine drinkers (high on both) and in non wine-drinkers (low on both), verbal overshadowing effects dissipate. However, when there is a mismatch in perceptual and verbal abilities, as there would be for untrained wine drinkers, verbal overshadowing effects are pronounced. Thus, consistent with a retrieval-based account, verbal overshadowing occurs when the contents of the verbalization fail to appropriately match the perceptual experience.

Further support for retrieval-based interference comes from a meta-analysis on the verbal description-identification relationship performed by Meissner et al. (2008). Their analysis, encompassing 33 articles and 4,278 participants, examined the association between five measures of face description quality and subsequent identification accuracy. The authors found that the number of errors that are generated in the description of a face is the strongest predictor, amongst the five measures, of whether the face will subsequently be accurately recognized ($r = -.18$, $p < .001$). Furthermore, results indicated that accuracy, not quantity, of details predicts performance on identification accuracy, and "congruence" between the description and the face identified were significantly related. Each of these findings suggests that the (in)accuracy of the contents of the descriptions is predictive of identification performance in support of a retrieval-based account.

Despite replications of findings whereby inaccuracies in verbal descriptions lead to inaccuracies in face identifications, retrieval-based interference cannot account for all verbal overshadowing findings. For instance, in a study conducted by Dodson, Johnson, and Schooler (1997), participants viewed a photo of both a man and a woman and later described either the male face or the female face. On a succeeding identification test they found that regardless of which face the participants verbalized, their recognition accuracy was inferior for both faces, compared to a control group that gave no verbal description. Additional studies have shown that describing a parent's face or even describing an unrelated stimulus such as a car (Westerman & Larson, 1997) can inhibit recognition of a target person's face. Thus, it becomes clear that if inaccuracies in a verbal description of a face are what account for inaccuracies in the target face recognition, then retrieval-based interference fails to explain why comparable discrepancies in identification accuracy occur when the target face itself was never verbalized.

Transfer-inappropriate processing shift (TIPS) theory provides an alternative, process-based, explanation of verbal overshadowing (Schooler, 2002). This account is based upon the premise that giving a verbal description of a face vs. recognizing a face requires two different cognitive processing orientations. More specifically, TIPS theory argues that when people provide a verbal description of a face they are using a “featural” processing orientation; however, when they are asked to recognize a face, a more non-verbal, “configural” processing orientation is used. Because it is the individual features of a face that are usually verbalized (and not the configural look) a shift in processing, from featural to configural, occurs when people change from verbalizing a face to trying to identify/recognize it. Schooler (2002) proposes that it is the brains inability to adequately transfer or shift memory from one processing orientation to the other (from featural to configural) that leads to the verbal overshadowing effect. This is congruent with “transfer appropriate processing” theory (Tulving & Thompson, 1973) which states that memory performance is strongest when the processing orientation at encoding matches the orientation necessary during retrieval. Additionally, Schooler (2002) proposes that as a witness transfers from one processing orientation to another, the original memory trace of the perpetrator’s face is unaltered, but becomes temporarily “unavailable” to make an accurate identification.

In contrast to a retrieval-based account, TIPS theory does not propose that the accuracy of the verbal description is directly related to the accuracy of the identification test. Instead, it is the act of providing a description, regardless of its accuracy, that will lead to an inappropriate processing shift and thereby verbal overshadowing on an identification test. Because of this, the processing account of TIPS theory offers a credible explanation for experiments conducted by Dodson et al. (1997) and Westerman and Larsen (1997) whereby verbalization of non-target faces or unrelated stimuli have been shown to create a verbal overshadowing effect.

Further support for the processing account of TIPS theory comes from a study by Macrae and Lewis (2002). In line with TIPS, the authors proposed that as long as a featural processing orientation is activated, an inhibition in later recognition or identification (configural processing) would occur. Accordingly, they believed that featural processing could be elicited in a number of ways (other than giving a verbal description) and that any task that elicits featural processing would inhibit subsequent identification accuracy. In their study participants were presented with a brief video depiction of a bank robbery, and then manipulated into either a "global" (configural) or "local" (featural) processing orientation. More specifically, in the processing orientation manipulation participants performed a letter-identification task, where they were presented a series of large letters composed of smaller mismatching letters (such as an H composed of Xs). In the "global" condition the task for participants was to identify the large letter, in the "local" condition they had to identify the small letters. Results of the study showed that, as predicted, compared to control condition participants, participants in the "local" processing orientation condition showed significant impairment in face recognition.

Taken together, two competing theories can account for verbal overshadowing findings. First, a retrieval-based interference account suggests that the depth of processing and variations in output criterion that participants use to retrieve their memory during the description task influences whether negative or positive effects are demonstrated (Meissner et. al, 2001). The (in)accuracy of the contents of a description can lead to a qualitative change in the memory representation. When inaccurate information is generated in the description a misinformed memory will be relied upon at retrieval and impair recognition accuracy. Second, transfer (in)appropriate processing suggests that when the cognitive processes at encoding do not match the processes relied upon at retrieval negative effects will occur (Schooler, 2002). However, when encoding processes match the retrieval processes positive effects will occur (McCrae & Lewis, 2002).

Verbal Facilitation

In the typical verbal facilitation paradigm, participants see a series of faces (e.g., 24 faces) which are presented for a brief period of time (e.g., 5s each). Following each face there is a brief interval during which participants are either asked to write a description of the face or complete a filler task. Following this interval the next face in the series is presented and the cycle of trials continues until all faces have been presented. Immediately following the presentation of the faces a standard yes/no recognition task is administered whereby the participants see the faces shown during the study phase intermixed with novel distractor faces. At this time, participants are asked to respond *yes* if the face they are shown appeared during the study phase and *no* if it did not appear. Recognition accuracy is assessed using signal detection measures of discrimination accuracy and response bias.

Using the aforementioned paradigm, Brown and Lloyd-Jones (2005) first demonstrated the verbal facilitation effect across a series of four studies. In their first study the authors simply manipulated the description task (description vs. no description control). Results demonstrated a significant main effect of descriptions on discrimination accuracy such that descriptions facilitated recognition accuracy relative to no descriptions. In the subsequent studies, Brown and Lloyd-Jones (2005) demonstrated that the verbal facilitation effect could be found regardless of the type of description task. Namely, whether the verbal description instructions called for describing similarities or differences between two sets of faces (study 3), or whether descriptions focused on featural aspects vs. holistic aspects of faces (study 4), verbal facilitation effects were found. This effect has also been replicated using both typical and distinctive faces (Brown & Lloyd-Jones, 2006), and own- and other-race faces (Susa, Meissner, & De Leon, in prep). In addition, a number of studies have also demonstrated verbal facilitation effects using paradigms that are more traditionally used in examining verbal overshadowing (Itoh, 2005; Meissner, Brigham, & Kelley, 2001; Yu & Geiselman, 1993).

In a recent meta-analysis on the verbal facilitation effect, based upon six research articles and 725 participants, Susa (in prep) found a significant, medium-sized effect of verbal descriptions enhancing discrimination accuracy (d = .55) of faces when compared to a no description condition. In addition, a small yet significant effect was found for verbal descriptions leading participants to be more likely to respond that they had seen the face before (liberal response bias; d = -.17). Overall, results from this analysis demonstrated that under the conditions of the verbal facilitation paradigm, verbal descriptions can have a reliable and robust effect on enhancing subsequent face recognition accuracy.

From a theoretical perspective, research on the cognitive processes involved in verbal facilitation has focused on a *levels-of-processing* approach. This approach can be separated into two distinct accounts, namely visual or semantic processing. Visual processing accounts claim that verbal facilitation arises as a function of the quantity or quality of information encoded about the face, such that descriptions enhance the quantity of facial features encoded which in turn facilitates recognition (Winograd, 1978). In contrast, semantic processing accounts suggest that descriptions provide strong semantic tags that are encoded with the face and consequently facilitate retrieval (Schooler, Ryan & Reader, 1996). Studies that have attempted to distinguish visual from semantic processing accounts have reached mixed conclusions and it is unclear at this time which theory best explains the verbal facilitation effect.

While retrieving memory in the form of a verbal description has been shown to have both negative and positive effects on subsequent face recognition memory, analogous negative and positive effects of retrieval on subsequent memory have been noted in the basic memory literature. It is believed that understanding the basic research and the theoretical underpinnings that have been proposed to account for repeated retrieval effects on memory can guide us in better understanding the cognitive mechanisms that might be responsible for explaining the negative vs. positive effects of descriptions on face memory.

How Retrieval Influences Memory

Research over the past several decades has demonstrated rather powerful effects of retrieval of information from memory on subsequent retrieval of the same information (repeated retrieval attempts). Bjork (1975) apply noted these powerful effects by stating that tests of repeated retrieval themselves should not be considered a neutral event, but rather a circumstance that modifies or changes memories and their future retention. For events that actually occurred and are retrieved, the act of retrieval makes them more likely to be recalled at a later time. However, for events that are retrieved erroneously the act of retrieval also consolidates the memory and makes it more likely to be erroneously retrieved later (Roediger, McDermott, & Goff, 1997). Therefore, much like the effect of verbal descriptions on subsequent face recognition, repeated retrieval attempts in other domains have also demonstrated paradoxical effects on memory such that repeated retrieval attempts can have either positive or negative effects (see Table 1). A discussion of the positive and negative effects of repeated retrieval attempts from the basic memory literature follows below.

Positive effects of retrieval. A long history of the positive effects of prior retrieval exists within the basic memory literature (see Roediger & Challis, 1989; Roediger, Wheeler, & Rajaram, 1993). In one of the more classic repeated retrieval studies, Ballard (1913) first demonstrated that repeated retrieval attempts could enhance later memory. He coined this phenomenon *reminiscence*, or "the remembering again of the forgotten without re-learning." In his study he gave school children passages of poetry to memorize and then had them make repeated retrieval attempts of the material. He found that the students would often recall more information in the latter retrieval attempts than they did in the earlier attempts. The positive effects of repeated retrieval have also been demonstrated by Erdelyi and Becker (1974) who demonstrated that recall of pictures increased over three repeated retrieval attempts. They subsequently coined the term, *hypermnesia*, to represent the overall increase in performance between repeated retrieval attempts over time.

In what is known as the *testing effect*, the positive effects of repeated retrieval have also been demonstrated in studies that "test" people's memory for learning material (Kang, McDermott, & Roediger, 2007; Roediger & Karpicke, 2006a; Spitzer, 1939). Specifically, students who are tested on material improve retention of the tested knowledge in comparison to students who are not tested, students who only study, or students who are only exposed to the test questions (see Roediger & Karpicke, 2006, for a review). In the typical testing effect paradigm participants are presented with material and told that they will be asked to later recall what the have learned. Immediately after the presentation of the material they are given either an initial test of the material, given time to restudy the material, or given a filler task to complete.

Finally, following a retention interval, they are administered a free recall assessment test asking them to remember the material that they were originally presented. While most of the experiments on the testing effect have been completed in a laboratory with word lists (McDaniel & Masson, 1985; Tulving, 1967) or picture lists (Wheeler & Roediger, 1992), the testing effect has also been found in educational settings using text materials (Roediger & Karpicke, 2006).

In one of the most recent testing effect studies Roediger and Karpicke (2006b) showed that repeated testing, in comparison to repeated studying, can improve retention after a one-week delayed assessment test. In their study participants read a prose passage and then either studied the passage once and took three intermediate free recall tests (STTT), or studied three times and take one intermediate free recall test (SSST), or studied the passage four times (SSSS). A two-minute filler task was given between testing/study periods. Students then took a free recall assessment test of the material one week later. As expected, after a one-week retention interval STTT participants recalled a significantly greater number of details than participants in the SSSS condition.

One of the keys to the success of repeated retrieval or testing seems to involve the act of self-generating a retrieval response. The *generation effect* is a robust finding demonstrating that information people self-generate is better remembered than material that is simply read (or is provided by the experimenter; Slamecka & Graf, 1978). In the typical experiment, participants are presented with a paired-associates list, most commonly word lists. Half of the pairs are provided intact by the experimenter and the participant is simply asked to read the pair (e.g., animal: horse). For the remaining items participants are presented with a cue word but must generate the target word from a fragment (e.g., animal: ho____). Subsequently, participants are asked to retrieve the associated word pairs from memory, with the basic finding that generated items are more readily retrieved than read items. Over the last two decades an extensive literature on the generation effect has developed, demonstrating the effect with different kinds of stimuli, including pictures (Kinjo & Snodgrass, 2000), math problems (Smith & Healy, 1998), and general knowledge questions (DeWinstanley, 1995). In addition, the effect appears to hold across different memory tasks including various forms of recall and recognition tests (Mulligan & Lozito, 2004). In the latest meta-analysis on the generation effect, Bertsch et al. (2007) synthesized primary research from 86 studies yielding 17,711 participants and 445 effect sizes. Results indicated that in comparison to participants who simply read a word, participants who generated a word demonstrated a memory benefit of almost a half a standard deviation (Cohen's d = .40).

Negative effects of retrieval. The experiments reviewed above have shown that repeated retrieval attempts of memory generally increase the likelihood of accurately retrieving the same memory at a later time. In particular, retrieval on a first attempt may make the memory more accessible later and even allow for the retrieval of additional, associated information. However, the effects of repeated retrieval are not always positive, as numerous studies have demonstrated that repeated retrieval can cause forgetting, interference, and false recollections (Roediger & Guynn, 1996; Roediger et al., 1997).

For example, Roediger and McDermott (1995; see also, Deese, 1959) demonstrated that repeated retrieval attempts can lead participants to falsely recall "lure" words from a studied list of associated words. In their experiment, participants were presented with lists of multiple words that were all associated with a critical non-presented word. For example, one list consisted of words such as, *bed, test, awake, tired, dream, wake,* and *snooze* all of which are associated with the non-presented word of *sleep*. Subsequently, participants either completed a recall test of the words that they were presented with or completed a filler task. When the recall test was administered participants falsely recalled the lures on 55% of the lists. After 16 lists had been studied, participants were then given a recognition tests of the studied words mixed with the lures and were ask to make an old (studied) or new (non-studied) decision about which words were originally presented to them. Results of the study indicated that false recognition rates approximately equaled those of veridical recognition and that the prior retrieval (recall) attempt increased overall recognition for both veridical and false responses. That is, prior retrieval lead to improved memory for words on the list and increased false recollection of a lure item that had never been presented.

With another classic paradigm, Loftus and colleagues (e.g. Loftus, Miller, & Burns, 1978) popularized decades of studies investigating the negative effects of retrieval in eyewitness memory when post-event misinformation is provided to witnesses. The typical paradigm involves participants viewing a series of slides depicting a crime, after which they are presented with a narrative or suggestive questions that contain mostly correct information but also *misinformation*. Participants then are given a cued recall test of each item that was presented and asked to determine if the source of their memory stems from the original slides or from the narrative. After a given delay (e.g. two days) participants return to the lab and are asked to take a second recall test and are instructed to recall only information that was presented in the original slides. Not surprisingly, generation of a false memory on the first recall attempt greatly increases the likelihood of false recall on later retrieval attempts, as well as the likelihood that participants say that they actually remember the occurrence of a misinformed item in the original slides. Since this original study demonstrated the negative effects of misinformed memory on later retrieval attempts numerous manipulations within the same standard paradigm have shown consistent results (for a review see, Gerrie, Garry, & Loftus, 2005), including a study by Loftus and Ketcham (1983) which demonstrated the negative retrieval effects of misinformed memory on subsequent recognition of faces.

As a predecessor to the study of Meissner et al. (2001), Roediger et al. (1993) assessed the impact of a forced recall procedure during an initial recall test on subsequent retrieval attempts. In their experiment participants studied 60 images of objects in the context of a story, and were told to remember as many of the items as possible for a later memory test. Thereafter, half of the participants were dismissed and asked to report back one week later, while the other half of the participants were given a recall test that involved either a free recall assessment (which warned them against guessing) or a forced recall test in which they were required to recall all 60 items from memory. A week later all participants returned to take either a free recall or a forced recall memory test (for those who took the initial test the format was kept the same). Results of the second retrieval attempt indicated that the number of correct items recalled between the free recall and forced recall participants was approximately the same; however, forced recall participants were more likely to self-generate items that were never presented. In line with the studies of Loftus et al., Roediger et al. concluded that forcing participants to self-generate misinformation has a powerful influence on source interference in later retrieval.

In recent years the effects of forced recall and confabulation have also been examined in the eyewitness memory paradigm. For instance, Chrobak and Zaragoza (2008) used a forced fabrication paradigm to show that people can develop false memories even when they are forced to fabricate an entire event that extends in time and involves people and places that they have never seen. Specifically, their results indicated that participants showed a minimal effect of forced fabrication on a one week recognition test; however, when participants returned 8 weeks later they freely reported their forced fabrication nearly 50% of the time - even when they had correctly rejected them on the 1-week recognition test. The authors concluded that participants were likely to freely report fabrication after 8 weeks because their source memory for having fabricated these events likely faded faster than their memory for the content of their fabrications (see also Ackil & Zaragoza, 1998).

While most studies on the generation effect have demonstrated positive effects on subsequent repeated retrieval attempts, Lane and Zaragoza (2007) have also demonstrated, in an eyewitness paradigm that the act of generating erroneous information can lead to negative repeated retrieval effects. Specifically, their studies examined whether encouraging eyewitnesses to generate descriptions of suggested items (fictitious) would increase subsequent false memories for these items in repeated retrieval attempts, relative to eyewitnesses who simply read descriptions of the items. Results of their study demonstrated that indeed this was the case, the act of generating perceptual details for a suggested item relative to reading the perceptual details, increased false memories on repeated retrieval attempts.

A final example of the negative influence of repeated retrieval returns our discussion to the *testing effect*. As previously noted, repeated testing of information often leads to positive effects on later memory retention; however, repeated testing can also have negative effects when an initial test provides participants with inaccurate information. Whereas teachers would never consider presenting students with erroneous information in their lectures, true-false and multiple-choice tests routinely expose students to misinformation. Each multiple choice question includes one correct answer and several false answers. For some items, students may select the wrong alternative, and because the act of retrieval increases the likelihood of subsequent retrieval, when they see that erroneous answer again on a later assessment they are more likely to believe that it is the correct answer. For instance, Toppino and Brochin (1989) gave students a true-false test and later asked them to judge the truth of objectively false statements to which they had been exposed, mixed with false items that they had not been exposed. Results of their study indicated that repeated false items were more likely to be judged as true than were novel false items.

A recent series of studies by Marsh, Roediger, and colleagues has also demonstrated the negative effects of repeated retrieval from multiple-choice testing. Most recently, Marsh, Agarwal, and Roediger (2009) demonstrated in three experiments that repeated multiple-choice tests involving SAT questions can have a negative influence on memory retention for low achieving students who do not score well on the initial retrieval attempt (test). Their data suggested that the negative effects of repeated retrieval arise from students selecting incorrect alternatives on the multiple-choice tests and believing that their answers are in fact correct (see also Roediger & Marsh, 2005).

Taken together, repeated retrieval attempts have both positive and negative effects on subsequent memory retention – much like the act of describing a face can have both positive and negative effects on subsequent recognition. In the next section, I examine the cognitive mechanisms that have been used to explain the effects of repeated retrieval in the basic memory literature. Interestingly, two general approaches to explaining these effects have emerged that are consistent with the two approaches described previously in the face literature. One principle, referred to as transfer appropriate processing (Tulving, & Thompson, 1973), is that performance on the final retrieval attempt will be most accurate when the cognitive processes used during the initial retrieval attempt match the processes used at the final retrieval attempt. The second principle involves a retrieval-based approach in which the content (accuracy) of the initial retrieval attempt governs subsequent retrieval accuracy. As will be seen, these two competing explanations mirror that of the TIPS and retrieval-based accounts in the verbal overshadowing vs. facilitation literatures discussed previously. An overview of how each of these accounts explains the positive vs. negative effects in the aforementioned findings will be discussed below. It is believed that each of the each account offers a credible explanation for the paradoxical effects; however, the purpose of the present study will be to further investigate the strengths and limitations of the retrieval-based account in hopes of better understanding the positive and negative effects that verbal descriptions can have on face memory.

Transfer appropriate processing. Several lines of research on the positive vs. negative effects of repeated retrieval attempts can be interpreted from the governing principles of transfer appropriate processing. In particular, transfer appropriate processing is often cited as the theoretical explanation for positive effects that result from testing and generation. The idea behind transfer-appropriate processing in accounting for the testing effect is that performance on a test of memory benefits to the extent that the processes required to perform well on the test matches the processes used during encoding. Thus, the same study strategies or processes of encoding that may aid performance on one type of test may have no effect on a different type of test that emphasizes a different type of processing. In most studies investigating the testing effect, transfer appropriate processing theory appears to hold as the match between cognitive processes engaged at initial retrieval and subsequent (repeated) retrieval enhances memory more than the process of studying the material. However, the transfer-appropriate processing account cannot explain why a match in test format does not always aid in performance, or why a mismatch in test format sometimes will aid performance. For example, Kang, McDermott, and Roediger (2007) examined this *mismatch effect* with short-answer and multiple-choice tests. Students in their experiment studied empirical research articles and then immediately took a short-answer or multiple-choice test. Three days later the students then took a final assessment test in either short-answer or multiple-choice format. Results indicated that the initial short-answer test produced the best retention for subsequent assessment tests in both formats (see also, Glover, 1989).

Transfer appropriate processing is also used to explain the positive effects of generation, in what McNamara and Healy (2000) refer to as a procedural account. The procedural account suggests that two key factors contribute to the generation effect. The first factor is that participants are engaging in cognitive processes that associate the target item to information that is already stored in memory. The second factor is that participants are able to reinstate the learning procedures that were used at study during the memory test. Consequently, the generation effect occurs because generating information is more likely than reading to facilitate cognitive processes during encoding that can be reinstated during the memory test – thereby enabling a match in the processes engaged at encoding vs. retrieval. A series of studies by Crutcher and Healy (1989) and McNamara and Healy (2000) used arithmetic problems to provide support for this theory, demonstrating that the generation effect only occurred when participants reinstated at test the cognitive procedures that were used during study (see also Bertsch et al., 2007; Steffens & Erdfelder, 1998).

Retrieval-based account. As an alternative to transfer appropriate processing, retrieval-based accounts have also been used to explain the positive vs. negative effects of retrieval. For example, an alternative explanation for the generation effect is the idea that generation requires more cognitive *effort* than reading which leads to greater test performance (McFarland, Frey, & Rhodes, 1980). Accordingly, the larger the effort required to process the target stimuli and thereby engage semantic processing of the information, the larger the generation effect size should be. The latest meta-analysis on the generation effect seems to lend support to this account as easier generation tasks produced smaller effects (*d* = .32) when compared with more difficult generation tasks (*d* = .55; Bertsch et al., 2007). Similarly, retrieval effort can also explain the positive effects of testing. Namely, recall tests require greater retrieval effort or depth of processing than recognition tests, and consequently have been shown to lead to stronger testing effects (Bjork, 1975; Kang et al., 2007). Experiment 1 will assess this prediction in the face memory domain.

Retrieval-based accounts also claim that the contents of the initial retrieval can change what is remembered during repeated retrieval attempts. This nicely accommodates many of the negative effects of repeated retrieval attempts associated with “forced” response manipulations (Chrobak & Zaragoza, 2008; Meissner et al., 2001; Pezdek et al. 2007; Roediger et al. 1993), misinformation from lures (Roediger & McDermott, 1995), suggestive questioning (Loftus et al., 1978; Loftus & Ketcham, 1983), and recollecting false alternatives from multiple-choice testing (Marsh, Agarwal, & Roediger, 2009; Marsh & Roediger 2005; Roediger & McDermott, 1995). Experiments 2 and 3 will assess the role of output criterion in yielding positive vs. negative effects on face recognition.

One manner in which to place the role of effort and veracity of recall into context is to consider the theoretical framework of monitoring and control processes offered by Koriat and Goldsmith (2002; see Pansky, Koriat, & Goldsmith, 2005, for a review). This framework is based on the assumption that people do not report everything that comes to mind, but rather they control their memory reporting in relation to their personal and situational goals. Two types of metacognitive control have been established. The first is a *report option*, which involves choosing whether or not to report information. The second regards *control over grain size*, which relates to the level of detail (precision) one might report information (Pansky et al., 2005). Koriat and Goldsmith (1996) have delineated between metacognitive monitoring and control processes. Monitoring processes assess the accuracy of the information that is retrieved from memory, while control processes govern the output of information based on an accuracy criterion. As such, output criterion is established on the basis of a trade-off between the gain of reporting accurate information vs. the cost of providing inaccurate information.

Koriat and Goldsmith (Goldsmith et al., 2002) argue that control processes mediate the effect of monitoring on the accuracy of information that is reported during a retrieval episode. In other words, it is the control processes that determine what to do with the monitored information before it is reported (or not). The accuracy of what one reports can be enhanced by a strict response criterion; however, this generally comes at the expense of reducing the quantity of accurate information provided. Similarly, control over grain size involves an accuracy vs. informativeness trade-off such that individuals must sometimes compromise between accuracy and informativeness when determining what information to report.

The ability to monitor information and regulate memory reporting is essential to determining retrieval performance. As such, reporting inaccurate information reflects not only a failure of memory processes, but also a failure to monitor the information that comes to mind (Pansky et al., 2005). Many memory failures stem from such source-monitoring errors (Johnson, Hastroudi, & Lindsay, 1993), and the success or failure of monitoring on initial retrieval episodes are thought to contribute to subsequent repeated retrieval effects. For example, many studies that invoke forced confabulation (Chrobak & Zaragoza, 2008; Meissner et al., 2001; Roediger et al. 1993) or misinformation from lures (Roediger & McDermott, 1995) trigger false recollections of information without an attempt to monitor the information that is being retrieved.

However, positive effects of repeated retrieval can be found in studies that ask participants to monitor their initial retrieval attempts. For example, when Meissner et al. (2001) "warned" participants to only generate verbal descriptors of a face that they knew to be correct they found descriptions enhanced (facilitated) subsequent face recognition. Similarly, Roediger et al. (1993) found that when participants monitored their memory through confidence ratings, they were more likely to recognize target items and less likely to falsely recognize intrusion items on subsequent repeated retrieval attempts. Experiments 2 and 3 will assess the role of monitoring at initial vs. subsequent retrieval stages in the context of manipulating output criterion.

Overview of Experiments

In the current experiments I investigated the cognitive mechanisms that underlie the verbal description-identification relationship to determine why verbal descriptions of a face sometimes overshadow and at other times facilitate subsequent face memory. Previous research has established that retrieval-based mechanisms may be responsible for these paradoxical findings. The effects of generation (or repeated testing), output (response) criterion, and metacognitive monitoring and control processes during initial and subsequent retrieval attempts have been shown to yield both positive and negative influences on subsequent attempts at memory retrieval. The current experiments focused on the role of these factors as a potential governing mechanism determining positive vs. negative effects of retrieval in explaining verbal facilitation vs. overshadowing of faces. It is proposed that verbal facilitation is a product of the effective generation and retrieval of an accurate description of a face that both enriches the memory of that face and provides a consistent source for subsequent retrieval. Three experiments addressed these predictions. Specifically, Experiment 1 assessed how the act of generation facilitates subsequent face memory, relative to simply reading descriptions of faces. Experiments

2 and 3 assessed whether manipulating output criterion leads to misinformed descriptions that produce negative effects on subsequent retrieval, and the extent to which monitoring of this misinformation (at initial vs. subsequent retrieval attempts) alleviates any impairment effects.

EXPERIMENT 1

Experiment 1 examined the effect the act of generation has on subsequent retrieval of face memory. The main goal of Experiment 1 was to assess whether generating a free recall description of a face facilitates identification performance, in comparison to when the description is simply read or when the face is not described. The effect of generation on source memory was also assessed by measuring participants' memory for the background color of the computer screen that was presented behind each face. Previous research, in other domains, has demonstrated positive effects of generation on repeated retrieval attempts when retrieval does not encourage suggestive misinformation (Bertsch et al., 2007; Lane & Zaragoza, 2007). It is believed that the mere act of generating a description requires more cognitive *effort* than reading a description, which in turn leads to greater semantic processing and subsequently to greater identification performance (McFarland, Frey, & Rhodes, 1980). Based on this research it was expected that generating a free recall verbal description of a face would facilitate identification performance relative to all other conditions.

Method

Participants

Ninety students (96% Mexican-American, 4% Mexican-National; 70% female) from the University of Texas at El Paso were recruited from Introduction to Psychology courses for participation in this study. This sample size was determined by conducting a power analysis, using an estimated effect size of d = .50, with alpha = .05 and power = .80. Average age of the participants was 20.24 years. All participants received course credit for their time.

Design

A one-way between-participants design was used to assess the act of generation (free recall vs. read vs. control) on face and source memory accuracy. Identification performance for the faces was measured through signal detection measures of hits, false alarms, discrimination accuracy (d'), and response bias (C). Source memory for the background color that was presented behind each face was measured using the Average Conditional Source Identification Measure (ACSIM; Vogt & Broder, 2007). This measure reflects the proportion of correct source claims made for each background color.

Materials

The experiment was conducted on PC computers running Medialab software. All instructions and images were displayed on 19" LCD monitors at 1280 x 1024 resolution. Forty eight Mexican-American faces of college-aged males were used as the experimental stimuli. Two photographs of each face were used (smiling vs. neutral facial expression). The faces were approximately 5" X 5" in size and featured only the face and neck of the person so that no other identifying features could be recognized. Participants in the read condition were provided with written descriptions of the faces they viewed. These descriptions consisted of five modal features of the face that were generated by 10 participants prior to the experiment under free recall conditions for a period of 60s.

Procedure

The paradigm used in Experiment 1 followed the general procedures used in verbal facilitation studies as developed by Brown and Lloyd-Jones (2005, 2006). When participants entered the lab they were greeted and asked to read an informed consent form. They were then seated at an individual computer station where a research assistant informed them regarding the general format of the experiment. The experiment began on the computer with the study phase. In the study phase, participants viewed 12 sequentially presented Mexican-American faces for a period of 5s each, which were preceded by a fixation cross presented for 250ms. Following the presentation of each face there was a 1s pattern mask. Faces were presented in random order, half of which were presented in front of a white background screen and half were presented in front of a black background screen (details provided below).

Prior to viewing the faces, participants were instructed to study each face for the entire time that it appeared on the screen. In addition, they were informed to remember the background color that was presented behind each face because they would be asked later in the experiment to recognize the faces they saw and whether they were presented in front of a white or black background. Following the presentation of each face there appeared a 1s pattern mask, followed by a 60s interval during which participants in the generation and read conditions either provided a description or read a description regarding what the face looked like. Participants in the free recall condition were asked to use the time to generate as many descriptors about the face as they could. The specific instructions read as follows: “Please be as complete in your description as possible so that another person seeing only your description could get as accurate an idea as possible of what the face is like.” Participants in the read condition were asked to use the time to read a description that someone else had written to about the face. Participants in the control condition were asked to use the 60s to complete anagram puzzles.

Immediately following the study phase, recognition of the faces and their source (background color) was tested in a yes/no decision task in which the pose of the study faces was changed from smiling to neutral and the background color behind each face was gray. 24 faces were presented to the participants, 12 of the faces were seen during the study phase and 12 were not. The faces were sequentially presented in random order. Participants were instructed to respond *yes-white* if the face appeared in the study phase and was presented with a white background, or *yes-black* if the face appeared in the study phase and was presented with a black background. If participants did not remember the face being presented in the study phase they were asked to respond, *no-new face*. Each face appeared in view until participants responded.

Results & Discussion

Experiment 1 examined the influence of the act of generation on subsequent retrieval of face memory. I hypothesized that generating a free recall description of a face would facilitate identification performance relative to participants who simply read a description or did not describe the face. Identification performance was initially separated into the proportion of correct identifications (hits) and the proportion of false identifications of faces (false alarms). These hit and false alarm rates were then used to compute signal detection estimates of discrimination accuracy (d') and response bias (c). Table 2 summarizes participants' performance across all four measures and the ACSIM measure of source memory. With regard to discrimination accuracy, a one-way between-participants ANOVA confirmed my predictions showing that there was a significant effect for the act of generation on face identification performance, $F(2,87) = 27.52$, $p < .001$, $\eta_p^2 = .39$. Pairwise comparisons indicated that participants who generated free recall descriptions demonstrated significantly greater discrimination accuracy for faces relative to participants who read the descriptions, $t(58) = 2.14$, $p = .037$, $d = .55$, or participants who completed the control task, $t(58) = 7.50$, $p < .001$, $d = 1.94$. In addition, participants who read the descriptions demonstrated significantly greater discrimination accuracy relative to control participants, $t(58) = 5.00$, $p < .001$, $d = 1.30$. No significant differences were observed with regard to response bias, $F(2, 87) = 2.72$, $p = .07$, $\eta^2 = .06$, or source memory, $F(2,87) = .75$, $p = .48$, $\eta_p^2 = .02$.

Overall, Experiment 1 demonstrated that participants who engaged in generation were significantly more accurate in identifying faces in comparison to participants who simply read the descriptions or who completed the anagram puzzles. These results are consistent with generation effects demonstrated in other domains of memory (see Bertsch et al., 2007, for a review) and suggest that verbalization of faces requires cognitive effort which enriches their semantic encoding and facilitates subsequent retrieval. Experiment 2 examined how the powerful of effects of generation on subsequent retrieval accuracy may be moderated by forcing participants to self-generate misinformation (cf. Chrobak & Zaragoza, 2008; Meissner et al., 2001; Pezdek et al., 2007; Roediger et al., 1993). In addition, Experiment 2 examined whether the negative effects of generating misinformation could be alleviated by providing participants memory training through feedback during the repeated retrieval (identification) attempt (Lane et al. 2007).

EXPERIMENT 2

Previous research has demonstrated that memory retrieval can have both positive and negative effects on subsequent (repeated) retrieval of the same information (Roediger et al., 1996). From a retrieval-based perspective it appears that the accuracy of the initial retrieval attempt likely moderates the accuracy of the subsequent retrieval attempt. For instance, Chrobak and Zaragoza (2007) demonstrated that forcing participants to confabulate details of an event can later lead them to believe details of the event that never actually occurred. Experiment 2 examined the effects of misinformation (during generation, via output criterion), and memory training through feedback (during repeated retrieval attempts) on face and source memory accuracy. The main purpose of Experiment 2 was to assess whether the paradox of the verbal description-identification relationship could similarly be explained by the accuracy of the initial retrieval attempt (the description). Specifically, does generating an inaccurate description lead to impairment effects on subsequent face identification? Experiment 2 also assessed whether providing memory training through feedback during the repeated retrieval attempt (identification) could alleviate this impairment (Lane et al., 2007).

One manner in which to vary the accuracy of information at retrieval is to manipulate output criterion (cf. Chrobak & Zaragoza, 2008; Meissner et al. 2001; Pezdek et al. 2007; Roediger et al. 1993). For participants who provided a description of the faces, half were "forced" to generate detailed descriptions containing confabulated information by being told to describe everything they could about the face even if they felt they were guessing. The other half generated shorter descriptions containing few, if any, confabulated details through "free" recall instructions.

If manipulating the criterion can mitigate the positive effects of generation, then how can we alleviate the problem? To answer this question in Experiment 2 memory training through feedback was implemented using procedures analogous to Lane et al. (2007, Exp. 3). In their experiment they demonstrated that it is possible to mitigate the influence of self-generated misinformation by providing positive feedback. Specifically, participants in their experiment had to determine if the information they remembered came from an event they had previously witnessed or from another source (i.e., misleading questions provided by the experimenter). At test (training phase) participants were either provided feedback or not regarding the accuracy of their source attributions. In a subsequent “assessment phase,” all participants were asked additional questions about what they remembered, absent further feedback being provided. Results of the Lane et al. (2007) study indicated that participants who were provided feedback during the training phase of the experiment had significantly greater accuracy in recall than those who were not provided feedback. The authors attributed the increased accuracy in the feedback condition to a greater ability to ascertain “online” the features of the information that were the most discriminative for distinguishing between sources.

Experiment 2 similarly provided memory training to participants (or not) through feedback on face and source memory decisions. The critical question regarded the extent to which memory training might improve face memory for participants in the forced recall condition. A main effect for output criterion was predicted such that participants in the free recall condition would demonstrate significantly greater face discrimination accuracy than participants in the forced recall and control conditions. A main effect for memory training through feedback was also predicted such that feedback would improve identification performance relative to those who do not receive feedback. I also predicted that feedback would have its strongest effect on participants in the forced recall condition such that it would alleviate any negative effects of forced confabulation (misinformation) on identification performance.

Method

Participants

Two-hundred and forty students (95% Mexican-American, 5% Mexican-National; 65% female) from the University of Texas at El Paso were recruited from an Introductory Psychology course for participation in this study. This sample size was determined by conducting a power analysis, using an estimated effect size of $d = .50$, with alpha = .05 and power = .80. Average age of the participants was 20.32 years. All participants received course credit for their time.

Design

A 3 x 2 between-participants design was used to assess the influence of output criterion (free recall vs. forced recall vs. control) and memory training (feedback vs. no feedback) on face identification performance. Identification performance for face and source memory were measured in the same way as Experiment 1.

Materials

All stimuli and presentation mediums replicate those employed in Experiment 1.

Procedure

The procedures employed here largely replicate those used in Experiment 1 for the free recall condition. One difference was that participants viewed 18 faces during the study phase instead of 12. Prior to viewing the faces, participants were instructed to study each face for the entire time that it appeared on the screen. In addition, they were informed that they would be asked later in the experiment to recognize the faces they saw and whether they were presented in front of a white or black background screen. Another critical difference in this experiment was that output criterion was manipulated for participants who described the faces. Specifically, one-third of participants were asked to generate free recall descriptions, while another one-third of participants were forced to recall descriptions that included inaccurate details. The remaining participants completed a control task.

Following the presentation of each face, a 1s pattern mask was presented followed by a 60s interval during which participants either generated a free or forced recall description of what each face looked liked. (Participants in the control condition solved anagram puzzles during this period of time.) Participants in the free recall condition were asked to freely recall accurate details about the face they just saw for a period of 60s. Specifically, prior to the study phase participants were provided with instructions notifying them that they would have to write a description of the faces in a 60s time interval. These instructions for free recall participants read: “Please be as complete in your description as possible, so that another person seeing only your description, could get as accurate an idea as possible of what the face is like.” Participants who were in the forced recall condition were “forced” to recall as many details as they could for a period of 60s even if they felt they were guessing. They were given the following instructions in addition to the instructions of the free recall, condition: “Prior research has demonstrated the importance of reporting everything that you can remember about the individual on the slide. Try not to leave out any details about the face even if you think they are not important. You must write as much information as you can, even if you start to feel that you are guessing. You should generate at least eight descriptive features.”

Immediately following the study phase, recognition of the studied faces and source was tested in a yes/no decision task in which the pose of the study faces was changed from smiling to neutral and the background screen was changed to gray. Participants were instructed to respond *yes-white* if the face appeared in the study phase with a white background and *yes-black* if the face appeared in the study phase with a black background. Participants were ask to respond *no-new face* if the face had not been presented previously. Each face appeared in view until the participant responded.

A key manipulation to the testing phase of the experiment was whether participants were provided memory training through feedback about correct face decisions they made. To examine this effect, 12 faces (six seen during the study phase, and six new) consisted of the "training portion" of the experiment, during which participants were either given memory training through feedback (feedback condition) or they completed a "practice" assessment without feedback (no feedback condition). Specifically, in the memory training portion of the test phase, trained participants were informed that following each face memory decision they would see the word "correct" or "incorrect" flash on the screen for a period of 2s to reflect the accuracy of their decision about whether the face was seen previously and whether it was presented in front of a white or black background. Furthermore, they were instructed that they should use the feedback to assess and improve the accuracy of their judgments for the faces that would follow. Participants in the no training condition did not receive feedback or any instructions pertaining to feedback, but also completed a 12 face "practice" phase of the experiment. In total, participants were trained on 12 faces during the training phase. Half of the faces were old faces (3 white background and 3 black background) and the remaining half were new faces that were not presented during study. The faces were presented sequentially and in random order.

Immediately following the "training" or "practice" portions of the test phase the "assessment" portion began. In the assessment portion, the remaining 24 faces (that were not used in the training or practice portion) were presented to the participants. Again, 12 of the faces were seen during the study phase and 12 were not. The faces were sequentially presented in random order. The instructions were identical for all participants, and feedback was not given. It took participants approximately 40 min. to complete the experiment.

Results & Discussion

Experiment 2 examined the effects of misinformation and memory training through feedback (during repeated retrieval attempts) on face memory accuracy. I hypothesized that participants who generated a misinformed (inaccurate) description through forced recall would demonstrate impaired identification performance relative to participants who gave a free recall description. Further, I hypothesized that this impairment could be alleviated using memory training through feedback during a training phase of the repeated retrieval (identification) task.

Description Quality. To measure the quality of the descriptions generated, two independent raters coded the descriptions for the number correct, incorrect, and subjective details produced from a random sample of 20 participants per condition, across 3 faces (with distinctiveness ratings at the mean and +/- 1 standard deviations from the mean). With these ratings I was able to compute the total number of details generated and the accuracy the descriptions generated (number of correct details, divided by the total number of details provided). Table 3 summarizes the results of description quality for Experiment 2. A 2 x 2 between-participants MANOVA was conducted to measure the effects of output criterion (free recall vs. forced recall) and memory training through feedback (feedback vs. no feedback) on description quality using the measures of the number of correct details, the number of incorrect details, and the number of subjective details produced. The results indicated a significant multivariate main effect for output criterion, $F(3,74) = 8.14$, $p < .001$, $\eta_p^2 = .25$. There were no significant differences with regard to feedback, $F(3,74) = 1.72$, $p = .171$, $\eta_p^2 = .07$, and no interaction effect, $F(3,74) = 1.26$, $p = .29$, $\eta_p^2 = .05$. With regard to output criterion univariate analysis indicated that forced recall participants, relative to free recall participants, produced more incorrect details, $t(78) = 3.44$, $p = .001$, $d = .77$, and more subjective details, $t(78) = 2.69$, $p = .009$, $d = .60$. However, no differences were found between forced and free recall participants in the number of correct details produced, $t(78) = 1.69$, $p = .095$, $d = .38$.

A 2 x 2 ANOVA was conducted to measure the effects of output criterion (free recall vs. forced recall) and memory training through feedback (feedback vs. no feedback) on both the total number of details produced and the accuracy of the details produced. With regard to output criterion a main effect was found for the total number of details produced such that participants who gave a forced recall description produced more overall details than participants who gave a free recall description, $t(78) = 4.32$, $p < .001$, $d = .96$. No significant differences were found on the accuracy of the details produced, $t(78) = 1.52$, $p = .133$, $d = .34$. No significant effects were found with regard to feedback, and no significant interaction effects were found. I also examined the correlations between each of the five measures of description quality and identification accuracy, however, none of the correlations proved to be significant.

Identification Accuracy. Similar to the first experiment, identification performance was initially separated into the proportion of correct identifications (hits) and the proportion of falsely identified faces (false alarms). These hit and false alarm rates were then used to compute signal detection estimates of discrimination accuracy (d') and response bias (c). Table 4 summarizes participants' performance across all four measures and the ACSIM measure of source memory.

A 3 x 2 between-participants ANOVA was conducted to measure output criterion (free recall vs. forced recall vs. control) and memory training through feedback (feedback vs. no feedback) on face memory performance. Results indicated that there was a significant main effect for output criterion on discrimination accuracy, $F(2,234) = 41.32$, $p < .001$, $\eta_p^2 = .26$. Pairwise comparisons indicated that participants who generated free recall descriptions demonstrated significantly greater discrimination accuracy for faces relative to participants who gave a forced recall description, $t(158) = 2.47$, $p = .015$, $d = .40$, or participants who completed the control task, $t(158) = 8.95$, $p < .001$, $d = 1.42$. In addition, participants who gave a forced recall description showed greater discrimination accuracy for faces relative to participants who completed the control task, $t(158) = 6.29$, $p < .001$, $d = .99$. Given the verbal facilitation nature of the paradigm it was not unexpected to find facilitative effects for participants in the forced recall condition. Most interesting is the finding that forced confabulation led to impairment in overall identification performance relative to giving a free recall description.

Results of Experiment 2 further indicated that there was no main effect of feedback on discrimination accuracy, $F(1,234) = 3.07$, $p = .081$, $\eta_p^2 = .01$, and no interaction effect between output criterion and memory training through feedback on discrimination accuracy, $F(2,234) = .865$, $p = .42$, $\eta_p^2 = .01$. Given predictions that memory training through feedback would moderate the effect of forced confabulation, I conducted several pairwise comparisons focused on the potential benefits of source memory training within each condition. As predicted, memory training did show a significant benefit for participants in the forced recall condition, $t(78) = 1.97$, $p = .05$, $d = .44$. However, no effects were seen the free recall or control conditions. It appears that the lack of an interaction was a product of low power ($1-\beta = .20$) necessary to obtain a modest interaction effect size. These findings are consistent with the research of Lane et al. (2007) and suggest that source memory training through feedback may be a practical way to enhance identification even when inaccurate descriptions are generated.

With regard to response bias, results indicated that there was a significant main effect of feedback on response bias such that participants who received feedback were more conservative (less likely to say that they had seen the face before) in their judgments relative to no feedback participants, $t(238) = -1.99$, $p = .047$, $d = .25$. There was no significant main effect of output criterion on response bias, $F(2,234) = .007$, $p = .993$, $\eta_p^2 = .00$, and no significant interaction between output criterion and memory training through feedback on response bias, $F(2,234) = .407$, $p = .667$, $\eta_p^2 = .00$. Also, there was no effect for either output criterion or memory training through feedback on source memory.

Overall, the results demonstrated that the verbal facilitation paradigm led to an improvement in face identification performance regardless of whether participants gave a free or forced recall description. However, the fact that there was a relative increase in performance for participants in the free recall relative to the forced recall is consistent with the research of Meissner et al. (2001) and Roediger et al. (1993), suggesting that generating forced confabulation makes people more likely to inaccurately recall details during repeated retrieval attempts. In support of a retrieval-based perspective, the present study demonstrated that forcing participants to give verbal descriptions led to an overall increase in the number of inaccurate details that were recalled and subsequently impaired their ability to make an accurate identification. Finding that memory training through feedback can increase identification performance when participants are forced to confabulate details suggests that feedback during subsequent retrieval attempts can alleviate negative effects produced during the initial retrieval. This finding is consistent with the research of Lane et al. (2007). The purpose of Experiment 3 was to demonstrate whether the impairment effects of forced confabulation could be even further strengthened if participants were asked to monitor the output of their initial retrieval.

EXPERIMENT 3

Studies by Koriat and Goldsmith (1996) and Roediger et al. (1993) have highlighted the importance of monitoring during an initial retrieval episode as a strong influence on subsequent retrieval attempts. For instance, an unpublished study by Roediger, Challis and Wheeler, cited by Roediger et al. (1993), demonstrated that confidence rating judgments that are made during an initial retrieval episode are predictive of performance during subsequent retrieval attempts. In their experiment participants studied 60 objects in the context of a story, and were told to remember as many of the objects as possible for a later memory test. After the 6 min. presentation of the story half of the participants were dismissed and asked to report back one week later, and the other half of the participants were given a recall test, which was either free recall, warning them against guessing, or a forced recall test. A week later all participants returned to take either a free recall or a forced recall memory test (for those who took the initial test the format was kept the same). An important detail of the procedure was that, after each item that participants produced, they were required to rate their confidence as to whether it was in the original story. For the purposes of the present experiment, the key finding of this study was that on the tests given one week later, target items that were rated with high confidence on the initial test were correctly recognized more often (and intrusion items rated with low confidence were correctly rejected more often) on the subsequent retrieval attempt, in comparison to the items seen by participants who did not take the initial test. This was true for both free recall and forced recall participants.

In Experiment 3 I assessed whether encouraging effective monitoring during an initial retrieval episode would serve to facilitate subsequent face memory performance. Namely, I proposed that effective monitoring during the initial retrieval would allow participants to effectively “tag” their confidence in each detail recalled and consequently control the accuracy of their subsequent memory retrieval attempts. In Experiment 3, output criterion (free recall vs. forced recall vs. control), and retrieval monitoring (confidence ratings vs. no ratings) at the time of initial retrieval (description) were manipulated. A significant interaction was predicated such that there would be no difference in identification performance between participants in the free recall condition regardless of whether they gave confidence ratings or not. However, in the forced recall condition participants who gave confidence ratings were predicted to have significantly higher identification performance than participants who did not give confidence ratings.

Method

Participants

Two-hundred students (92% Mexican-American, 8% Mexican-National; 64% female) from the University of Texas at El Paso were recruited from an Introductory Psychology course for participation in this study. This sample size was determined by conducting a power analysis, using an estimated effect size of d = .50, with alpha = .05 and power = .80. Average age of the participants was 19.26 years. All participants received course credit for their time.

Design

A 2 x 2 + 1 between participants design was used to measure output criterion (free recall vs. forced recall), and retrieval monitoring (confidence ratings vs. no ratings) with a hanging control on face identification performance. Identification performance was assessed in the same way as Experiments 1 and 2.

Materials

All stimuli and presentation media replicate those employed in Experiment 2.

Procedure

The procedures employed here largely replicate those used in Experiment 2 such that participants' output criterion was manipulated as they described 18 faces during the study phase. The critical difference was that half of the participants were asked to monitor their initial retrieval by providing confidence estimates for each feature they described. The other half of participants did not monitor their initial retrieval (cf. Roediger et al. 1993). Participants in the retrieval monitoring condition were asked to rate their confidence in the accuracy of each of the features they described immediately after writing each feature (on a Likert-type scale 1 = not at all confident, 7 = very confident). At the end of the 60s interval the next fixation cross was presented and the process was repeated. Following the study phase, participants' recognition of studied faces and their source memory was assessed consistent with the no-training feedback condition procedures employed in Experiment 2. That is, participants were provided with an initial practice phase involving six old and six new stimuli absent any performance feedback, followed by an assessment phase involving the remaining 12 old and 12 new stimuli. It took participants about 40 min to complete Experiment 3.

Results and Discussion

Experiment 3 examined the effects of output criterion (free recall vs. forced recall vs. control) and retrieval monitoring (confidence ratings vs. no ratings) on face memory accuracy. The output criterion manipulation was consistent with the procedures employed in Experiment 2 and was used to force participants to confabulate details in their descriptions. The key distinction in this study was that participants were no longer given feedback during the subsequent retrieval attempt but were rather asked to make confidence ratings (or not) about the accuracy each descriptive feature that they generated during the description task. I hypothesized that retrieval monitoring would reduce the number of inaccurate features that participants would generate even when they were forced to generate as much information as they could. In addition, I hypothesized that by reducing the number of incorrect details generated the negative effects of forced recall on identification performance would be equal to the accuracy level of participants in the free recall conditions.

Description Quality. Description quality was measured in the same way as Experiment 2. Table 5 summarizes the results for Experiment 3. A 2 x 2 between-participants MANOVA was conducted to measure the effects of output criterion (free recall vs. forced recall) and retrieval monitoring (confidence ratings vs. no ratings) on description quality using the measures of the number of correct features, the number of incorrect features, and the number of subjective features. The results indicated there was a significant multivariate main effect for output criterion, $F(3,74) = 4.35$, $p = .007$, $\eta_p^2 = .15$. There was also a significant multivariate main effect for retrieval monitoring, $F(3,74) = 4.52$, $p = .006$, $\eta_p^2 = .16$. There was no significant multivariate interaction effect, $F(3,74) = 1.91$, $p = .14$, $\eta_p^2 = .07$. With regard to output criterion univariate analysis indicated that forced recall participants produced more incorrect details than free recall participants, $t(78) = 2.80$, $p = .006$, $d = .63$. However, no differences were found between forced and free recall participants in the number of correct, $t(78) = 1.09$, $p = .277$, $d = .25$, or subjective details, $t(78) = 1.07$, $p = .288$, $d = .24$, produced. With regard to retrieval monitoring, univariate analyses indicated that participants who gave confidence ratings produced fewer incorrect details than participants who did not give confidence ratings, $t(78) = 2.29$, $p = .025$, $d = .52$. However, no differences were seen in the number of correct details produced, $t(78) = 1.95$, $p = .055$, $d = .44$, or in the number of subjective details produced, $t(78) = 1.70$, $p = .093$, $d = .38$. Given predictions that retrieval monitoring would moderate the effect of forced confabulation on incorrect details, I conducted univariate analyses for the interaction of output criterion and retrieval monitoring on the measure of incorrect details. As predicted, there was a significant interaction for the number of incorrect details produced, $F(1,76) = 3.80$, $p = .05$, $\eta_p^2 = .05$. Pairwise comparisons demonstrated that retrieval monitoring reduced the number of incorrect details for participants in forced recall condition, $t(78) = 2.99$, $p = .005$, $d = .68$. However, no effects of retrieval monitoring were seen between participants in the free recall conditions, $t(78) = .358$, $p = .722$, $d = .08$.

A 2 x 2 ANOVA was conducted to measure the effects of output criterion (free recall vs. forced recall) and retrieval monitoring (confidence ratings vs. no ratings) on both the total number of details generated and on the accuracy of the details generated. A significant main effect was found for both output criterion and retrieval monitoring on the total number of details generated. With regard to output criterion, the results suggested that participants who gave a free recall description generated less details than participants who gave a forced recall description, $t(78) = 3.17$, $p = .002$, $d = .70$. With regard to retrieval monitoring, participants who gave confidence ratings generated significantly less details than participants who did not give ratings, $t(78) = 2.79$, $p = .007$, $d = .62$. No significant effects were found on description accuracy.

Identification Accuracy. Identification performance was measured using estimates of discrimination accuracy (d') and response bias (c). Similar to Experiments 1 and 2, identification performance was initially separated into the proportion of correct identifications (hits) and the proportion of falsely identified faces (false alarms). These hit and false alarm rates were then used to compute signal detection estimates of discrimination accuracy (d') and response bias (c). Table 6 summarizes participants' performance across all four measures.

A 2 x 2 between-participants ANOVA was conducted to measure output criterion (free recall vs. forced recall) and retrieval monitoring (confidence ratings vs. no ratings) with a hanging control on face identification performance. With regard to discrimination accuracy, results indicated that there was a significant main effect for output criterion on face identification performance, $F(1,156) = 4.57$, $p = .034$, $\eta^2 = .03$. Pairwise comparisons indicated that participants who generated free recall descriptions demonstrated significantly greater discrimination accuracy for faces relative to participants who gave a forced recall description, $t(158) = 2.11$, $p = .037$, $d = .34$, or participants who completed the control task, $t(158) = 6.23$, $p < .001$, $d = 1.20$. In addition, participants who gave a forced recall description showed greater discrimination accuracy for faces relative to participants who completed the control task, $t(158) = 4.31$, $p < .001$, $d = .83$. There was no main effect of retrieval monitoring on discrimination accuracy. However, the main effect of output criterion was qualified by a significant interaction effect, $F(1,156) = 5.83$, $p = .017$, $\eta^2 = .03$. Pairwise comparisons indicated no significant differences in discrimination accuracy for participants in the free recall condition regardless of whether they gave confidence ratings or not, $t(78) = 1.28$, $p = .20$, $d = .28$. However, participants in the forced recall condition who gave confidence ratings showed significantly greater identification performance than those who did not give ratings, $t(78) = 2.12$, $p = .037$, $d = .47$. No significant effects were found with regard to response bias, or source memory accuracy.

Overall, the results demonstrating the main effect of output criterion on discrimination accuracy were consistent with the results from Experiment 2. The manipulation of retrieval monitoring in Experiment 3 produced results that were consistent with my hypotheses and the research of Koriat and Goldsmith (1996) and Roediger et al. (1993). Namely, the results demonstrated that retrieval monitoring through confidence ratings during the initial retrieval attempt can significantly reduce inaccurate retrieval output and subsequently increase the accuracy of the repeated retrieval attempt. Much like the effect of output criterion, the effects of retrieval monitoring support a retrieval-based theoretical perspective.

GENERAL DISCUSSION

The current experiments investigated the cognitive mechanisms that underlie the verbal description-identification relationship and determined why verbal descriptions of a face sometimes overshadow and at other times facilitate subsequent face memory. Using a retrieval-based theoretical framework, I proposed that the effects of generation (or repeated testing), output (response) criterion, and metacognitive monitoring and control processes during initial and subsequent retrieval attempts would govern the positive vs. negative effects of retrieval in face memory. Three experiments addressed these predictions. Experiment 1 assessed how the act of generation facilitates subsequent face memory, relative to simply reading descriptions of faces. Results confirmed the predictions showing that people who generate verbal descriptions of faces are subsequently more likely to accurately identify each face in comparison to people who read a description or complete a non-description control task.

Experiment 2 assessed whether the powerful effects of generation could impair identification accuracy if the initial retrieval contained misinformation. Based on studies conducted by Meissner et al. (2001), Pezdek et al. (2007), and Roediger et al. (1993) it was believed that forcing participants to confabulate details in their descriptions would interfere with the accuracy of the original encoding of the face and subsequently impair identification accuracy. Results indicated that people who are forced to confabulate details, relative to people who gave a free recall description, produced more inaccurate details in their descriptions and subsequently were also less likely to correctly identify the faces. Further, the results indicated that memory training through feedback could alleviate the negative influence of forced confabulation, such that, even when people are forced to confabulate details memory training through feedback can enhance their accuracy in identification to levels equivalent to people who give free recall descriptions.

Experiment 3 assessed whether the negative effects of forced confabulation could also be alleviated through retrieval monitoring during the description task. According to research by Koriat and Goldsmith (1996) and Roediger et al. (1993), when people monitor the accuracy of information at an initial retrieval episode it reduces the level of incorrect information that is produced. In Experiment 3 people monitored the accuracy of their descriptions by giving confidence ratings for the accuracy of each descriptive feature generated. With regard to output criterion, the analysis of description quality was consistent with that of Experiment 2, showing that people who gave forced recall descriptions generated more incorrect details than people who gave free recall descriptions. With regard to retrieval monitoring, people in the forced recall condition who gave confidence ratings produced fewer incorrect details than people in the forced recall condition who did not give confidence ratings. The main effects for the number of incorrect details generated were qualified by a significant interaction such that there were no differences between those who gave confidence ratings and those who did not in the free recall condition, but, as indicated, those who gave ratings generated significantly less incorrect details in the forced condition. Identification accuracy results mirrored those of description quality. It appears that retrieval monitoring has a strong effect on moderating identification accuracy, as confidence ratings enhanced identification accuracy for people in the forced recall condition relative to those who did not give confidence ratings. In short, like memory training through feedback in Experiment 2, in Experiment 3 retrieval monitoring was found to alleviate the negative effects of forced confabulation. In contrast to Experiment 2, however, in Experiment 3 retrieval monitoring reduced the number of incorrect details produced during the description task.

The results of the current experiments provide strong support for a retrieval-based theoretical perspective on repeated retrieval attempts of memory. Much like other domains of memory, it appears that the positive vs. negative effects of the verbal description-identification relationship are largely a product of the effects of generation (or repeated testing), output (response) criterion, and metacognitive monitoring and control processes. Verbal facilitation resulted as a product of an accurate description generation, and was moderated when metacognitive monitoring and control processes were being used during both initial and repeated retrieval attempts of memory recall. Further, the decrease in identification performance for forced recall participants demonstrated that when initial retrieval involves inaccurate information the negative effects of generation for face memory are similar the negative effects of selecting incorrect lures in multiple-choice testing (e.g. Marsh et al. 2009). Namely, generating inaccurate information once makes it more likely that people will falsely recognize that information as being accurate during repeated retrieval attempts of memory. With regard to face memory, identification performance declines as a result of the self-generated inaccurate information interfering with the original memory trace.

The results of the current experiments support a generation effect theoretical perspective of verbal facilitation which suggests that the mere act of generation enhances cognitive effort and depth of semantic encoding. Specifically, the data suggest that when a person describes a face the mere act of generation encodes the information at a depth that makes retrieval more accessible. Given that there is no delay in the verbal facilitation paradigm between the encoding of the face and the description task, it makes intuitive sense that the description would strengthen the encoding of the face. However, this would likely dissipate in a paradigm such as verbal overshadowing where there is a considerable delay between the encoding and description tasks. Future research should examine the how the act of generation is moderated by the delay between tasks, while accuracy of the description is controlled.

From a cognitive processing standpoint, the data also suggest that participants are able to adjust their retrieval processing through feedback to accurately ascertain which faces had been previous seen. In particular, participants who confabulate details appear to be using feedback to calibrate their decisions better. It is not entirely clear at this time how people use feedback to enhance their decision-making abilities, other than to say that feedback allows participants to discover for themselves the diagnostic characteristics of faces that improve identification performance (Lane et al., 2007). Future research needs to determine both the diagnostic features of face identification and the metacognitive strategies that feedback elicits to enhance performance. In particular I think it is important to ascertain whether feedback is enhancing decisions through intuitive vs. deliberative processing. Is it possible that people are consciously using feedback to discern the diagnostic features of accurate source information, if so can we train people during retrieval to monitor these diagnostic features properly without providing feedback? It is also important to note that memory training through feedback did little to enhance source memory as source performance scores were at chance levels. It is believed that the measuring source through the verbal facilitation paradigm was simply too difficult a task, resulting in a floor effect.

The results of the Experiment 3 support Koriat and Goldsmith's (1996) model of retrieval monitoring, showing that when people monitor the accuracy of recall they will control their output to reduce the recall of inaccurate details. The data suggest that by having to report confidence ratings for the accuracy of the details that were generated, participants in Experiment 3 used control processes to mediate the effect of monitoring on the accuracy of information that was reported. Precisely, in comparison to participants who did not report confidence ratings, those that did set a more strict response criterion, which was essential in determining retrieval performance. The effects of retrieval monitoring on identification performance appear to be strong. Relative to feedback, which occurred during subsequent retrieval attempts, monitoring during the initial retrieval, strengthened the encoding of the original memory trace and more effectively enhanced identification performance for participants in the forced recall condition. As such, the results of Experiment 3 suggest that an accurate retrieval of information during both initial (description) and subsequent (identification) retrieval attempts involves an ability to effectively monitor information that comes to mind.

I believe that the application of these results is worthy of further research and exploration. Most importantly the results of the current studies suggest that when eyewitnesses are interviewed extreme caution should be taken to reduce the amount of incorrect information that is generated. Directions that warn participants about the dangers of providing incorrect information might be one practical way to reduce memory error (see Meissner et al., 2001), however this usually comes at the expense of also reducing the total amount of information that is provided which may be necessary to apprehend guilty perpetrators. Future research needs to continue to examine the utility of the accuracy of information provided and the best interviewing techniques to elicit the necessary ratio of accurate to inaccurate information. Further, the results of the current suggest that generation of a verbal description helps to strengthen the encoding of a face when the description immediately follows an eyewitness viewing the face. In order to reduce inaccurate descriptions and to strengthen encoding of the face every effort should be made to gather a description immediately after a crime has occurred. The current studies also suggest that providing witnesses with feedback training during identification procedures may be another way to enhance identification accuracy. There are many practical issues that make providing feedback problematic. If the investigator were to give incorrect feedback, the suggestiveness of the inaccurate feedback could further damage eyewitnesses' memory and their subsequent ability to correctly identify the correct perpetrator (Lane et al., 2007). I believe that future research needs to investigate ways in which proper feedback procedures can be properly implemented. Probably the most practical way that the results of the current study can inform interviewing techniques is through retrieval monitoring. The data suggest that if people rate their confidence in accuracy of each descriptive feature they generate, they will be less likely to report incorrect information, and subsequently more likely to make a correct identification. Future research should investigate the practical utility of eliciting confidence ratings from eyewitness descriptions during investigation procedures. One factor that might mitigate the effect of retrieval monitoring is social pressure. In an actual criminal investigation people may try to please the investigator by inflating their confidence ratings. Whether these inflated ratings would negate the benefits of monitoring needs to be empirically tested.

In conclusion, the current studies suggest that a retrieval-based theoretical perspective of memory can provide us valuable insight as to why verbal descriptions of faces at times impair and at other times facilitate our ability to correctly identify faces. As is the case in other domains of memory, the verbal description-identification relationship in face memory demonstrates the powerful effects of generation on subsequent memory retrieval attempts. Verbal facilitation is a product of the effective generation and retrieval of an accurate description of a face which can be enhanced through retrieval monitoring and feedback during the initial description and subsequent identification tasks. In contrast, the negative effects of retrieval result from the generation of incorrect details and a failure of effective monitoring at the time of initial retrieval. This leads to a memory that is believed to have high validity but thereby distorts or interferes with subsequent retrieval of the original memory trace. While eyewitnesses will continue to play a vital role in a criminal investigation, it is with a better understanding of the cognitive processes that underlie the verbal description-identification relationship that future research and law enforcement practices can move forward.

REFERENCES

Ackil, J. K., & Zaragoza, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology, 34,* 1358-1372.

Ballard, P. B. (1913). Oblivescence and reminiscence. *British Journal of Psychology, 1,* 1-82.

Bertsch, S. P., Pesta, B. J., Wiscott, R., McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35,* 201-210.

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

Brandimonte, M. A., Hitch, G. J., Bishop, D. V. M.(1992). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 157-165.

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

Brown, C., & Lloyd-Jones, T. J. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Memory & Cognition*, *34*, 277-286.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift Accounts of verbal overshadowing. *European Journal of Cognitive Psychology, 20 (3),* 396-413.

Chrobak, Q. M., & Zaragoza, M. S. (2008). Inventing Stories: Forcing witnesses to fabricate entire ficticious events leads to freely reported false memories. *Psychonomic Bulletin & Review, 15*, 1190-1195

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 669-675.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22.

DeWinstanley, P. A. (1995). A generation effect can be found during naturalistic learning. *Psychonomic Bulletin & Review, 2,* 538-541.

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition*, *25*, 129-139.

Erdelyi, M. H., & Becker, J. (1974). Hypernesia for pictures: Incremental memory for pictures but not for words in multiple recall trials. *Cognitive Psychology, 6,* 159-171.

Fallshore, M. & Schooler, J.W. (1995). The verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1608-1623.

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.

Gerrie, M., Garry, M., & Loftus, E. (2005). False Memories. *Psychology and law: An empirical perspective* (pp. 222-253). New York, NY US: Guilford Press.

Goldsmith, M., Koriat, K., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: General, 131*, 73-95.

Itoh, Y. (2005). The facilitating effect of verbalization on the recognition memory of incidentally learned faces. *Applied Cognitive Psychology, 19*, 421-433.

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

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

Kinjo, H., & Snodgrass, J. G. (2000). Does the generation effect occur for pictures? *American Journal of Psychology, 113,* 95-121.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490-517.

Lane, S. M., Roussel, C. C., Villa, D. & Morita, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source-monitoring errors. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33,* 1131-1142.

Lane, S. M., Zaragoza, M. S. (2007). A little elaboration goes a long way: The role of generation In eyewitness suggestibility. *Memory & Cognition, 35,* 1255-1266.

Loftus, E. F., & Ketcham, K. E. (1983). The malleability of eyewitness accounts. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 159-171). New York: Wiley.

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.

Macrae, C.N. & Lewis, H.L. (2002). Do I know you?: Processing orientation and face Recognition. *Psychological Science, 13,* 194-196.

Marsh, E. J., Agarwal, P. K., & Roediger, H. L. III. (2009). Memorial consequences of answering SAT II questions. *Journal of Experimental Psychology: Applied, 15,* 1-11.

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

McFarland, C. E., Frey, T. J., & Rhodes, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior, 19,* 210-225.

McNamara, D. S., & Healy, A. F., (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory & Language, 43*, 652-679.

Meissner, C.A. & Brigham, J.C. (2001a). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15*, 603-616.

Meissner, C.A., Brigham, J.C., & Kelley, C.M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29*, 176-186.

Meissner, C. A., Sporer, S. L., & Schooler, J. F. (2006). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology, Vol. 2*. Mahwah, NJ: Lawrence Erlbaum.

Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008) A theoretical and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology, 20 (3)*, 414-455.

Melcher, J. M., & Schooler, J. W. (1996). The misrememberance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory & Language, 35*, 231-245.

Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 175-214). San Diego, CA: Elsevier Academic Press.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness recall and testimony. In Brewer, N. & Kipling, W. D. (Eds.), *Psychology & Law: An empirical perspective* (pp.93-150). New York, NY US: Guilford Press.

Pezdek, K. Sperry. K., Owens. S. M. (2007). Interviewing Witnesses: The Effect of Forced Confabulation on Event Memory. *Law and Human Behavior, 27,* 463-478.

Ryan, R. S. & Schooler, J. W. (1998). Whom do words hurt?: Individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology*, *12*, 105-126.

Roediger, H. L. III, & Challis, B. H. (1989). Hypernesia: Increased recall with repeated tests. In C. Izawa (Ed.), Current issues in cognitive processes: The Tulane Floweree Symposium on Cognition (pp. 175-199). Hillsdale, NJ: Erlbaum.

Roediger, H. L., III, & Guynn, M. (1996). Retrieval processes. *Memory* (pp. 197-236). San Diego, CA: Academic Press.

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

Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1(3),* 181-210.

Roediger, H.L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. Journal of Experimental Psychology: Learning, Memory, & Cognition, 31, 1155-1159.

Roediger, H. L. III, McDermott, K. B. (1995). Creating false memories: Remembering words not Presented in lists. *Journal of Experimental Psychology: Learning, Memory & Cognition, 21,* 803-814.

Roediger, H. L. III, McDermott, K. B., & Goff, L.M. (1997). Recovery of true and false memories: Paradoxical effects of repeated testing. In M. A. Conway (Ed.), *Recovered memories and false memories* (pp. 118-149). Oxford: Oxford University Press.

Roediger, H. L. III, Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing and reconstructing the past. In D.L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 30, pp. 97-134). New York: Academic Press.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology*, *16*, 989-997.

Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*, 36-71.

Schooler, J. W., Ryan, R. S., & Reder, L. (1996). The costs and benefits of verbally rehearsing Memory for faces. In D. J. Herrmann, C. McEvoy, C. Hertzog, P. Hertel, & M. K. Johnson (Eds.) *Basic and applied memory research: Vol. 2. Practical applications* (pp. 51-65). Mahwah, NJ: Erlbaum.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4,* 592-604.

Smith, R. W., & Healy, A. F. (1998). The time-course of the generation effect. *Memory & Cognition, 26,* 135-142.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30,* 641-656.

Steffens, M. C., & Erdfelder, E. (1998). Determinants of positive and negative generation effects in free recall. *Journal of Experimental Psychology A: Human Experimental Psychology, 51A,* 705-733.

Susa, K. J. (in prep). A meta-analysis of the verbal facilitation effect in face recognition.

Susa, K. J., Meissner, C. A., & De Leon, D. (in prep). Verbal facilitation effects for own- and other-race faces.

Toppino, T. C., & Brochin, H. A. (1989). Learning from tests: The case of true-false Examinations. *Journal of Educational Research, 83,* 119-124.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 1,* 321-334.

Tulving, E., & Thompson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

Vogt, V., & Broder, A. (2007). Independent retrieval of source dimensions: An extension of Results by Starns and Hicks (2005) and a comment on the ACSIM Measure. *Journal of Experimental Psychology, 33,* 443-450.

Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3,* 240-245.

Winograd, E. (1978). Encoding operations which facilitate memory across the lifespan. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of memory* (pp. 252-262). San Diego: Academic Press.

Westerman, D.L., & Larsen, J.D. (1997). Verbal-overshadowing effect: Evidence for a general shift in processing. *American Journal of Psychology, 110,* 117-128.

Wilson, T.D., & Schooler, J.W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, *60(2)*, 181-192.

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

Table 1

Positive and Negative Effects of Repeated Retrieval Attempts

Positive	Negative
Testing Effect: People tested on material improve retention relative to those who were not tested	**Testing Effect:** When people select the wrong answer it increases the likelihood of accessing the wrong answer during subsequent retrieval
Generation Effect: Information that people generate is better remembered than information that is read	**Generation Effect:** When people are "forced" to generate inaccurate information it can lead to false memories that are easily accessed during repeated retrieval attempts of memory

Table 2

Descriptive Statistics for Identification Performance in Experiment 1

	Free Recall			Read			Control		
	M	(*SD*)	95% CI	*M*	(*SD*)	95% CI	*M*	(*SD*)	95% CI
Hits	.77	(.11)	[.73, .81]	.69	(.17)	[.63, .75]	.63	(.14)	[.58, .68]
False Alarms	.04	(.08)	[.01, .07]	.07	(.11)	[.03, .11]	.20	(.19)	[.13, .27]
Disc. Accuracy (*d′*)	2.76	(.70)	[2.51, 3.01]	2.36	(.76)	[2.09, 2.63]	1.42	(.68)	[1.18, 1.66]
Response Bias (*C*)	.59	(.36)	[.46, .72]	.64	(.54)	[.45, .83]	.36	(.55)	[.16, .56]
ACSIM	.57	(.18)	[.51, .63]	.53	(.18)	[.47, .59]	.59	(.19)	[.52, .66]

Note. CI = confidence interval.

Table 3

Descriptive Statistics for Description Quality in Experiment 2

	Free Recall			Forced Recall		
	M	(*SD*)	95% CI	*M*	(*SD*)	95% CI
No Feedback						
Correct	4.14	(.80)	[3.89, 4.39]	4.46	(.92)	[4.17, 4.75]
Incorrect	1.50	(.58)	[1.32, 1.68]	2.36	(.80)	[2.11, 2.61]
Subjective	.16	(.16)	[.11, .21]	.24	(.26)	[.16, .32]
Total	5.80	(1.00)	[5.49, 6.11]	7.06	(1.13)	[6.71, 7.41]
Accuracy	.72	(.09)	[.69, .75]	.67	(.09)	[.64, .70]
Feedback						
Correct	4.53	(.94)	[4.24, 4.82]	4.89	(.86)	[4.62, 5.16]
Incorrect	1.63	(.88)	[1.36, 1.90]	1.93	(.68)	[1.72, 2.14]
Subjective	.11	(.12)	[.07, .15]	.32	(.36)	[.21, .43]
Total	6.23	(1.20)	[5.86, 6.60]	7.14	(1.16)	[6.78, 7.50]
Accuracy	.74	(.12)	[.70, .78]	.73	(.07)	[.71, .75]

Note. CI = confidence interval.

Table 4

Descriptive Statistics for Identification Performance in Experiment 2

	Free Recall			Forced Recall			Control		
	M	(*SD*)	95% CI	*M*	(*SD*)	95% CI	*M*	(*SD*)	95% CI
No Feedback									
Hits	.68	(.16)	[.63, 73]	.64	(.17)	[.59, .69]	.53	(.19)	[.47, .59]
False Alarms	.08	(.11)	[.05, .11]	.10	(.09)	[.07, .13]	.15	(.15)	[.10, .20]
Disc. Accuracy (*d'*)	2.26	(.78)	[2.02, 2.50]	1.90	(.68)	[1.69, 2.11]	1.38	(.76)	[1.14, 1.62]
Response Bias (*C*)	.60	(.49)	[.45, .75]	.55	(.49)	[.40, .70]	.62	(.54)	[.45, .79]
ACSIM	.45	(.19)	[.39, .51]	.52	(.23)	[.45, .59]	.53	(.25)	[.45, .61]
Feedback									
Hits	.74	(.13)	[.70, .78]	.70	(.14)	[.66, .74]	.60	(.15)	[.55, .65]
False Alarms	.07	(.07)	[.05, .09]	.09	(.10)	[.06, .12]	.16	(.12)	[.12, .20]
Disc. Accuracy (*d'*)	2.39	(.55)	[2.22, 2.56]	2.20	(.68)	[1.99, 2.41]	1.40	(.51)	[1.24, 1.56]
Response Bias (*C*)	.46	(.44)	[.32, .60]	.50	(.49)	[.35, .65]	.43	(.44)	[.29, .57]
ACSIM	.47	(.17)	[.42, .52]	.52	(.17)	[.47, .57]	.51	(.24)	[.44, .58]

Note. CI = confidence interval.

Table 5

Descriptive Statistics for Description Quality in Experiment 3

	Free Recall			Forced Recall		
	M	(*SD*)	95% CI	*M*	(*SD*)	95% CI
No Ratings						
Correct	4.70	(.58)	[4.52, 4.88]	4.79	(1.02)	[4.47, 5.11]
Incorrect	1.50	(.63)	[1.30, 1.70]	2.24	(.81)	[1.99, 2.49]
Subjective	.16	(.20)	[.10, .22]	.14	(.24)	[.07, .21]
Total	6.36	(1.07)	[6.03, 6.69]	7.17	(1.19)	[6.80, 7.54]
Accuracy	.75	(.06)	[.73, .77]	.71	(.09)	[.68, .74]
Ratings						
Correct	4.24	(.72)	[4.02, 4.46]	4.55	(.84)	[4.29, 4.81]
Incorrect	1.43	(.68)	[1.22, 1.64]	1.58	(.58)	[1.4, 1.76]
Subjective	.18	(.21)	[.11, .25]	.31	(.30)	[.22, 40]
Total	5.84	(.74)	[5.61, 6.07]	6.43	(.71)	[6.21, 6.65]
Accuracy	.74	(.11)	[.71, .77]	.75	(.10)	[.72, .78]

Note. CI = confidence interval.

Table 6

Descriptive Statistics for Identification Performance in Experiment 3

	Free Recall			Forced Recall			Control		
	M	(*SD*)	95% CI	*M*	(*SD*)	95% CI	*M*	(*SD*)	95% CI
No Ratings									
Hits	.71	(.15)	[.66, .76]	.62	(.12)	[.58, .66]	.56	(.19)	[.50, .62]
False Alarms	.06	(.09)	[.03, .09]	.09	(.12)	[.05, .13]	.16	(.19)	[.10, .22]
Disc. Accuracy (*d′*)	2.49	(.80)	[2.24, 2.74]	1.97	(.76)	[1.73, 2.21]	1.51	(.75)	[1.28, 1.74]
Response Bias (*C*)	.57	(.49)	[.42, .72]	.65	(.40)	[.53, .77]	.57	(.66)	[.37, .77]
ACSIM	.50	(.16)	[.45, .55]	.54	(.19)	[.48, .60]	.58	(.26)	[.50, .66]
Ratings									
Hits	.67	(.16)	[.62, .72]	.69	(.18)	[.63, .75]			
False Alarms	.06	(.09)	[.03, .09]	.08	(.11)	[.05, .11]			
Disc. Accuracy (*d′*)	2.29	(.63)	[2.09, 2.49]	2.32	(.74)	[2.09, 2.55]			
Response Bias (*C*)	.64	(.50)	[.49, .79]	.62	(.52)	[.46, .78]			
ACSIM	.52	(.22)	[.45, .59]	.51	(.16)	[.46, .56]			

Note. CI = confidence interval.

CURRICULUM VITA

Kyle Joseph Susa was born on December 9, 1978 in Marshfield, Wisconsin. He is the son of Lee and Mary Susa, the brother of Jon and Matt Susa, the brother-in-law of Gusty and Angie Susa, and the uncle of Maya, Melanie, Mark, and Cooper Susa. He graduated from Lincoln High School in Wisconsin Rapids, Wisconsin in 1997. Upon graduation from high school, he attended the University of Wisconsin-Madison. Kyle graduated from the University of Wisconsin-Madison in May 2002 with a Bachelor of Science degree in Psychology and Secondary Education. After college, Kyle began teaching Psychology at Pius XI High School in Milwaukee, Wisconsin. During his time in Milwaukee he was also a Psychology Research Assistant for Dr. Anthony Greene at the University of Wisconsin-Milwaukee. In August of 2005, Kyle entered the Psychology Ph.D. program (Legal Psychology concentration) at the University of Texas at El Paso (UTEP). At UTEP, Kyle worked under the supervision of Dr. Christian A. Meissner, Dr. Roy S. Malpass, and Dr. Osvaldo Morera. Kyle's research interests include understanding the Social and Cognitive processes involved Eyewitness Memory and Face Recognition. In August of 2006 he began a three-year face recognition research fellowship with the U.S. Department of Homeland Security. Kyle has several research publications in the top-tier journals of Cognitive and Social Psychology and he has also given many conference presentations throughout North and Central America. In addition to his major research focus, Kyle also has an interest in research methodology and statistical analyses that apply to predicting player and team performance in sports.

Permanent address: 2611 Sampson Street

Wisconsin Rapids, WI 54494