2019

The Influence of Recall Instructions on the Verbal Overshadowing Effect

Melissa Ann Baker
University of Mississippi

THE INFLUENCE OF RECALL INSTRUCTIONS ON THE VERBAL OVERSHADOWING EFFECT

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctorate of Philosophy
in Experimental Psychology
in the Department of Psychology
The University of Mississippi

by

MELISSA A. BAKER

May 2019

ABSTRACT

The current study investigated the influence of recall instructions on the verbal overshadowing effect (VOE). Participants watched a video of a burglar breaking and entering, were asked to recall information about the burglar, and attempted to identify the burglar from a photo lineup. Recall instructions were varied between participants. In Experiment 1, participants in a 'general recall' instruction condition were instructed to provide a description of the burglar's physical appearance. Participants in a 'face recall' instruction condition were asked to provide a description of the burglar's facial features. A control group of participants, the 'no recall' instruction condition, were not asked to provide a description. Participants in the 'face recall' instruction condition demonstrated verbal overshadowing, while participants who were given 'general recall' instructions did not. Experiment 2 investigated the influence of recall instruction type (general, face) and recall instruction length (short, long) on the VOE. In Experiment 2, participants in a 'general short recall' instruction condition were given brief instructions to describe the perpetrator's physical appearance; in the 'general long recall' instruction condition, participants were given longer general recall instructions. Participants in the 'face short recall' instruction condition were given brief instructions to describe the perpetrator's face; participants in the 'face long recall' instruction condition were given longer face recall instructions. Participants in a control 'no recall' instruction condition were not given recall instructions. Participants in both the 'face recall' and 'general recall' instruction conditions demonstrated verbal overshadowing. Recall instruction length did not influence identification accuracy. Across

both experiments, the relationship of recall instructions to description accuracy and lineup identification response time was also assessed. Additionally, the relationships between description accuracy and identification accuracy as well as identification response time and identification accuracy were examined. While recall instructions appeared to influence measures of both identification accuracy and description accuracy, a meaningful relationship between description accuracy and identification accuracy was not found in the present study. Regarding lineup identification response times, the relationship between these times and identification accuracy is not clear. Overall, results have theoretical implications regarding the VOE and applied implications for how law enforcement officers administer lineups.
Keywords: verbal overshadowing effect, recall instructions, eyewitness memory, recall, recognition

DEDICATED TO

one old dog, Suey, who was no help at all in writing this dissertation.

LIST OF ABBREVIATIONS AND SYMBOLS

VOE	Verbal overshadowing effect
RBI	Retrieval-based interference
TIPS	Transfer inappropriate processing shift
CS	Criterion shift
FA	False alarm
d’	Discriminability
C	Response criteria
RT	Response time
OR	Odds ratio
Min	Minimum
Max	Maximum

ACKNOWLEDGMENTS

Thank you to my committee members, Dr. Matthew Reysen, Dr. Elicia Lair, Dr. Mervin Matthew, and Dr. Kimberly Kaiser, for their time and efforts spent reviewing this dissertation. I thank them for supporting this dissertation and providing thoughtful feedback always aimed at moving me forward.

I am especially thankful to Dr. Matthew Reysen for being a great mentor. I am so grateful for the opportunity to work in his research lab and learn from him.

I also acknowledge the many undergraduate research assistants for all of their time spent collecting data for the experiments included in this dissertation.

Lastly, I would like to thank my wonderful and enormous family for their love and support throughout the completion of this dissertation as well as my doctoral studies.

TABLE OF CONTENTS

LIST OF TABLES

CHATPER 1: INTRODUCTION

In 1990, Schooler and Engstler-Schooler reported the first demonstration of verbal overshadowing. They had participants watch a video of a perpetrator committing a crime. Prior to attempting to identify the perpetrator in a photo lineup, one group of participants was asked to verbally describe the perpetrator's face while other participants were not asked to provide a verbal description of the perpetrator. Schooler and Engstler-Schooler (1990) showed that providing a verbal description of a perpetrator's face, prior to attempting to identify the face, hindered later memory performance on a lineup identification task. This effect was labeled the verbal overshadowing effect (VOE) and was defined as fewer 'hits,' or instances of correct identifications, on the lineup task after providing a verbal description of a perpetrator.

A number of studies have replicated the VOE (e.g., Alogna et al., 2014; Brown & Lloyd-Jones, 2003; Dodson, Johnson, & Schooler, 1997; Smith & Flowe, 2014) although there is a notable exception (e.g., Clifford, 2003). Additionally, the large size of the effect reported by Schooler and Engstler-Schooler has been questioned (Francis, 2012; Meissner & Brigham, 2001; Schooler, 2011). More recently, Alogna et al. (2014) published a replication of Schooler and Engstler-Schooler's original study. They found a large VOE effect size consistent with the effect size reported by Schooler and Engstler-Schooler. Taken as a whole, the existing research indicates that verbal overshadowing is a legitimate phenomenon associated with a relatively large effect size. *One aim of the present study was to replicate the VOE using novel stimuli with a particular emphasis on the observed effect size.*

VOE Theories

Although the VOE has been observed in a number of situations, theoretical accounts of the phenomenon have been less successful in determining, definitively, which underlying cognitive mechanisms lead to impaired recognition performance following verbalization of a previously encoded perpetrator. The theoretical accounts that have received empirical support are the retrieval-based interference (RBI) account (Meissner, Brigham, & Kelley, 2001), the transfer inappropriate processing shift (TIPS) account (Schooler, 2002), and the response criterion shift (CS) account (Clare & Lewandowsky, 2004; Smith & Flowe, 2014).

Retrieval-based interference (RBI) account. According to the RBI account, generating inaccurate information, while verbally describing a face, can interfere with a subsequent lineup task (Meissner et al., 2001; Schooler & Engstler-Schooler, 1990). Thus, recall instructions that increase the probability of a participant recalling inaccurate information during the verbalization task (i.e., asking participants to recall every detail about an event) would be expected to increase the extent to which verbal overshadowing is observed: that is, poor quality descriptions should be correlated with poor identification accuracy performance. Consistent with the RBI account, previous research has shown that descriptions containing more incorrect details tend to be associated with less accurate recognition performance (e.g., Finger & Pezdek, 1999; Meissner et al., 2001). However, such a correlation has not always been observed, and these situations prove problematic for the RBI account (e.g., Brown & Lloyd-Jones, 2003; Schooler & Engstler-Schooler, 1990).

Transfer inappropriate processing shift (TIPS) account. Alternatively, the TIPS account suggests that the incongruence between the type of processing used to encode a face (configural processing) and the sequential nature of recalling information about the face during the verbalization task (featural processing, see Diamond & Carey, 1986) leads to reduced performance during recognition (Schooler, 2002). According to the TIPS account, the greater the discrepancy between the type of processing that occurs while encoding the face and the type of processing invoked during the verbalization task, the greater the predicted VOE. While researchers have found support for the TIPS account (e.g., Meissner & Brigham, 2001; Schooler, 2002), this theory has proven difficult to test and more research designed to test the TIPS account is needed.

Response criterion shift (CS) account. The response CS account is based on the assumption that the act of recalling a previously encoded face makes participants' response criteria, during the subsequent lineup identification task, more conservative. Thus, the CS account suggests that participants are less likely to identify both 'guilty' and 'innocent' faces from a lineup when they have previously attempted recall features of the face. For instance, Clare and Lewandowsky (2004) showed that participants who provided a description of a target's face, compared to participants who did not provide a description, were more likely to say the target was "not present" in the lineup identification task (i.e., they adopted a more conservative response criterion). The reason for this decrease in response criteria is not entirely clear, but Clare and Lewandowsky suggested that because participants find describing a face difficult, they tend to think that their memory for the face is poor, and become reluctant to choose a target from the lineup. This account, however, does not explain why describing a face can interfere with recognition when participants are forced to choose someone from a lineup (e.g., Dodson et al., 1997; Fallshore & Schooler; Schooler and Engstler-Schooler, 1990).

Clearly, current accounts of the VOE are insufficient to account for all of the accumulated data. As such, as noted by Mickes and Wixted (2015), the field may benefit from the development of new or revised theoretical accounts. The present studies might contribute to future theory development regarding the VOE.

Mechanisms driving the VOE

In addition to the various theories put forth to explain the VOE, there are two basic mechanisms by which providing descriptions might interfere with later identification accuracy. First, describing a previously seen face might interfere with a participant's ability to distinguish between a previously viewed face and other faces in the lineup. In other words, a participant's discriminability might be reduced. The RBI and TIPS accounts are consistent with this view, as those theories suggest that verbal overshadowing occurs because the act of recalling a face interferes with the original memory of the face, thus, impairing performance during recognition (Meissner et al., 2001; Schooler, 2002). Another possible explanation for the VOE is that providing a description reduces one's tendency to make an identification from a lineup. That is, a participant's response criteria might become more conservative. The response CS account suggests that providing a description of a face reduces one's inclination to choose a face from any type of lineup (Clare & Lewandowsky, 2004) and is, thus, consistent with this view.

A recent paper by Mickes and Wixted (2015), written in response to Alogna et al.'s (2014) replication report, noted the importance of identifying whether reduced discriminability or engendering a more conservative response bias has the greater influence on the VOE. As Mickes and Wixted note, both Schooler and Engstler-Schooler and Alogna et al.'s studies used only target present lineups. As a result, only the number of correct identifications (hit rates) were reported and the observed decrease in the hit rates between conditions was used to measure the

VOE. However, Mickes and Wixted argued that in order to determine the relative contributions of a decrease in discriminability versus inducing a more conservative response criteria, the proportion of innocent faces identified from a target absent lineup (false alarm rates) is needed, in addition to the hit rate. In fact, knowing the hit rate alone provides little information regarding whether VOEs are a result of reduced discriminability or a more conservative response criterion. *Another aim of this study was to address the question raised by Mickes and Wixted concerning the extent to which verbal overshadowing influenced participants' ability to discriminate between a perpetrator and an innocent suspect (discriminability) and their tendency to make an identification from a lineup (response criteria). In the present study, both target present and target absent lineups were used to obtain the necessary hit and false alarm rates needed to examine the influence of discriminability and response criteria on verbal overshadowing.*

Recall Instructions

The various VOE theories are often tested in studies in which several critical variables have been manipulated. One of the most important of these variables is the recall instructions that participants are given after viewing the video of the crime but prior to the identification task (e.g., Finger & Pezdek, 1999; Meissner, 2002; Meissner et al., 2001; Smith & Flowe, 2014). By definition, the VOE is a demonstration of impaired identification accuracy for participants who provide a description of a previously encoded face compared to participants who do not provide a description of a previously encoded face. However, research regarding the relationship between recall instructions and identification accuracy suggests that different types of recall instructions might result in different VOE sizes. For instance, Smith and Flowe (2014) showed participants a video of a theft, asked participants to describe the thief's face, and then asked them to identify the thief from a photo lineup. The researchers manipulated recall instructions: some participants were asked to provide a long and detailed description of the thief's face while other participants were given shorter less detailed description instruction of the face. The remaining participants were not asked to describe the thief's face. Smith and Flowe (2014) found that participants who were required to provide a long and detailed description of the thief's face had poorer identification accuracy compared to participants who were given shorter face description instructions and to the no recall participants. Similarly, Meissner et al. (2001) found poorer identification accuracy when participants were asked to provide a very detailed and complete description of a person's face compared to participants asked to provide a general description of the face or to participants who were not given recall instructions. Other researchers have found similar relationships between recall instructions and identification accuracy (e.g., Finger & Pezdek, 1999; Meissner, 2002).

Overall, it is clear that type of recall instructions can have an influence on the size of VOEs. Unlike previous studies, a condition in which participants were asked to describe a person freely or generally was included in the present study. That is, whether asking participants to provide a general description of a person influences verbal overshadowing differently than asking participants to provide a specific and detailed description of a person's face was investigated. Previous studies on recall instructions and identification accuracy have only included instructions solely focusing on the description of a person's face (e.g., Meissner et al., 2001; Schooler & Engstler-Schooler, 1990; Smith & Flowe, 2014). To my knowledge, there is no published research comparing VOEs of participants who provide general descriptions of a suspect's appearance versus participants who provide specific descriptions of the suspect's face.

The addition, in the present study, of a general recall condition offers an opportunity to test implications of both the TIPS and the RBI theories. For instance, if participants who provide a general description of the burglar have better identification accuracy compared to participants who provide a detailed description of the face, results would tend to support the TIPS account (e.g., Meissner & Brigham, 2001; Schooler, 2002). Additionally, if participants who describe the facial features of the person have poorer identification accuracy compared to participants who provide a general description of the burglar, and if participants who describe the facial features of the person have more incorrect details in their description compared to participants who provide a general description, then results would also support the RBI account (e.g., Finger & Pezdek, 1999; Meissner at al., 2001). In either case, results from a study that includes a general recall condition might have implications for the existing theories of the VOE.

An aim of this study was to examine identification accuracy when participants were asked to provide a general description of a previously seen burglar. Identification accuracy for participants asked to provide a general description was compared to identification accuracy for participants who were asked to provide a specific description of the face of a previously seen burglar. Identification accuracy of those two conditions was also compared to a control no recall instruction condition in which participants did not provide a description. I expected to find verbal overshadowing between participants asked to provide a detailed description of a suspect's face and participants who did not recall. However, because previous research has not compared identification accuracy between participants asked to provide a general description of a suspect to participants asked to provide a detailed description of a suspect's face, predictions about how recall instructions might influence identification accuracy were not clear.

Examining the difference in VOEs between people asked to describe a suspect in a general way compared to people asked to describe the facial features of a suspect also has important implications for procedures used by law enforcement officers when they question eyewitnesses about a crime. In these situations, instructions to eyewitnesses might be tuned in a way that promotes subsequent identification accuracy if and when the eyewitness is given a lineup identification task. At the moment, there is no standard procedure for law enforcement to follow when asking eyewitnesses to describe a previously seen suspect. There is a written set of suggested guidelines for law enforcement to follow when asking eyewitnesses to recall the physical description of a suspect. These guidelines written by the National Institute of Justice suggest that law enforcement should "…encourage the witness to report *all* [my italics] details, even if they seem trivial" (*Eyewitness Evidence: A Guide for Law Enforcement*, 1999, p. 22). Asking an "all" question might or might not promote identification accuracy. The results of the present studies might be used to address this question.

Description Accuracy and Identification Accuracy Relationship

The relationship between description accuracy and identification accuracy was also examined in the current study. Many studies using eyewitness paradigms have examined the relationship between description accuracy and identification accuracy for perpetrators. It seems intuitive that an eyewitness who provides an accurate description of a suspect would be more likely to do well on a lineup identification task than a less accurate eyewitness. In fact, the U.S. Supreme Court has openly endorsed the belief that there is a meaningful and useful relationship between identification accuracy and description accuracy (*Neil v. Biggers*, 1972). However, research seems to indicate that people who are superior at describing details of faces from memory are not equally superior at recognizing faces (Goldstein, Johnson, & Chance, 1979; Howells, 1938; Wolfskiel & Brigham, 1985).

In an early review, Sporer (1996) noted that although descriptions and identifications seem to be unrelated, they are nevertheless influenced by many of the same variables. For instance, both recall and recognition memory may be influenced by various factors at the time of encoding (e.g., weapons focus effect, stress, etc.) and retrieval (e.g., misleading questions, misinformation, etc.; Meissner, 2002). Recall instructions have also been found to have a similar influence on both recall and recognition memory. For instance, Meissner (2002) found a relationship between recall instructions and description accuracy as well as a relationship between recall instructions and identification accuracy. However, when Meissner examined the overall relationship between description accuracy and identification accuracy, no relationship was found. Other studies investigating the relationship between description accuracy and identification accuracy, while manipulating recall instructions, observed a similar lack of relationship between description accuracy and identification accuracy (e.g., Smith & Flowe, 2014).

It is puzzling that recall instructions can influence both description accuracy and identification accuracy, yet description accuracy might not be related to identification accuracy. This type of finding would suggest that the effects of recall instructions on identification accuracy are not being mediated by recall performance, and would be inconsistent with the RBI theory of verbal overshadowing. *An aim of this study was to examine the relationship between description accuracy and identification accuracy. The relationship between description accuracy and identification accuracy separately for each of the recall instruction conditions was also examined.*

Lineup Identification Response Time

Schooler and Englster-Schooler (1990, Exp 6) showed participants a photo of a face and later asked those participants to identify that face from a lineup. Before the lineup task, some participants were asked to describe the face and some participants were not. During the lineup task, some participants were forced to make a lineup decision within 5 s while some were given an unlimited amount of time to make a lineup decision. The researchers found an interaction between recall instructions and decision time: the recall task did not influence participants who were required to respond within the 5 s response limit. However, participants who provided a description of a face and who were given an unlimited amount of time to make a lineup decision had poorer identification accuracy compared to participants who did not provide a description of the face and who were given an unlimited amount of time to make a lineup identification decision.

Schooler and Engstler-Schooler’s (1990, Exp 6) findings suggests that the time it takes for an eyewitness to make an identification decision may be related to verbal overshadowing. While the amount of time participants had to make a lineup identification decision was not manipulated in the present study, identification response times were recorded in the present studies. Previous research on lineup identification response time and identification accuracy suggests that faster lineup identification response times are associated with more correct identifications from lineups, while slower response times are associated with more incorrect identifications (Brewer, et al., 2006; Bruer & Pozzulo, 2014; Dunning & Perretta, 2002; Sauerland & Sporer, 2009; Sporer, 1993; Weber et al., 2004). *An aim of the present study was to examine the amount of time it took participants to make their lineup identification decision as a function of recall instruction condition. The relationship between lineup identification response time and identification accuracy for each of the recall instruction conditions was also examined.* Consistent with previous research, faster lineup identification response times were expected to be associated with accurate lineup identifications. However, specific predictions about the relationship between recall instructions and lineup response times were not made. Differences found in the relationship between lineup identification response time and identification accuracy for each recall instruction condition would be informative for future theory development regarding the VOE. At the moment, research on lineup identification response time and identification accuracy has not been applied to research on the VOE; however, lineup response time data might offer valuable insight regarding the VOE and why it occurs.

The Current Study

The present study describes two experiments. Across two experiments, the study had five aims: The first was to replicate Schooler and Engstler-Schooler's (1990) VOE and examine the size of the VOE. Replications of the VOE have been successful, but the reported effect sizes have been inconsistent. An additional goal was to examine verbal overshadowing using original stimuli—a crime video and photo lineups that had not been used previously. A second aim was to address the question raised by Mickes and Wixted (2015) concerning the extent to which verbal overshadowing influenced participants' ability to discriminate between a perpetrator and an innocent suspect (discriminability) and their inclination to choose a suspect from a lineup (response criteria). The current study used both target present and target absent lineups so discriminability and response criteria could be properly examined. A third aim of the present study was to examine identification accuracy when participants were provided with certain recall instructions. In Experiment 1, participants in a 'general recall' instruction condition were instructed to provide a description of the burglar's physical appearance. Participants in a 'face recall' instruction condition were asked to provide a description of the burglar's facial features. A control group of participants, the 'no recall' instruction condition, were not asked for recall. In Experiment 2, participants in a 'general short recall' instruction condition were given brief instructions to describe the perpetrator's physical appearance; in the 'general long recall' instruction condition, participants were given longer general recall instructions. Participants in the 'face short recall' instruction condition were given brief instructions to describe the perpetrator's face; participants in the 'face long recall' instruction condition were given longer face recall instructions. Participants in a control 'no recall' instruction condition were not given a recall task. A fourth aim of the study was to examine the relationship between description accuracy and identification accuracy. A final aim of the study was to investigate the amount of time it took participants to make their lineup identification decision as a function of recall instruction condition. That is, the relationship between lineup identification response time and identification accuracy was examined.

CHAPTER II: EXPERIMENT 1

In Experiment 1, participants watched a brief video of a breaking and entering, were asked to recall the physical description of the burglar, and attempted to identify the burglar from a photo lineup. Recall instructions were varied between participants. There was a 'general recall' instruction condition in which participants provided a general description of the perpetrator that they saw in the video, a 'face recall' instruction condition in which participants provided a detailed description the perpetrator's facial features, and a 'no recall' instruction condition in which participants did not describe the perpetrator.

Method

Design

The study was a 3 (recall instructions: general recall, face recall, no recall) × 2 (lineup type: target present, target absent) between-groups factorial design. All participants were randomly assigned to conditions.

Participants

The participants included a sample of 212 undergraduate students (68.90% female, overall M_{age} = 18.47). All of the participants received partial course credit for participating. The study was administered using Qualtrics online surveying software. The study was approved by the University of Mississippi's Institutional Review Board.

Materials

Crime video. A 30 s video was recorded of a female criminal breaking into and burglarizing a home. The video was filmed in color. In the video, the burglar was shown entering a glass-paneled door, checking the surroundings, opening cabinet doors, placing stolen items into a bag, and briefly gazing toward the camera (ostensibly a surveillance system) prior to leaving the crime scene (screen shots of the video are in Appendix A).

Recall task. After viewing the crime video participants engaged in one of three recall instruction conditions. Participants in the 'general recall' instruction condition were told, "Please describe the person you saw in the video." Participants in the 'face recall' instruction condition were told, "Please describe the face you saw in the video. Your task is to describe the person in such a way that your description would aid someone else in attempting to identify the person. Your description should focus on facial features. Write about the shape and size of the eyes, eyebrows, nose, ears, mouth, chin, etc. Try not to leave out any details about the face even if you think they are not important." The 'face recall' instruction was based on the instructions used by Smith and Flowe (2014; Exp 1, 'standard recall' condition instructions). Participants in the 'general recall' and 'face recall' instruction conditions were provided 2 min to complete their descriptions. Participants typed their descriptions using a keyboard. Participants in the 'no recall' instruction condition played a game of Tetris for 2 min. All stimuli were presented, and all responses were recorded using laboratory computers.

In an effort to describe participants' ability to recall details about the perpetrator as accurately as possible, a total of four measures of description accuracy were computed. Corresponding with Meissner et al. (2001), *correct descriptors, incorrect descriptors, subjective descriptors*, and the *number of descriptors* were calculated for each participant. Correct descriptors were those that correctly described the perpetrator (e.g., race/ethnicity, gender, age, hair color, hair length, eye color, etc.). Incorrect descriptors were those that did not correctly describe the perpetrator (e.g., wearing glasses, freckles, tattoos, etc.). Subjective descriptors were those that were ambiguous, such as relative descriptors or descriptors about personality (e.g., 'she looked sneaky', 'average height', etc.). The number of descriptors included the total number of adjectives, or descriptors, in a participant's description. The number of descriptors was the sum, or total number, of correct, incorrect, and subjective descriptors regarding the perpetrator.

Photo lineups. The participants were either shown a target present or target absent simultaneous lineup. All of the lineups were presented in color and comprised 6 photos (a 2 × 3 display) that matched the verbal description of the perpetrator. This description was provided by a sample of 20 pilot participants (all 20 rated the female criminal). Those participants also viewed the initial lineup and provided feedback about any particular features of the persons or differences in backgrounds that could potentially influence identifications. That feedback was used to minimize bias in the lineups. All of the photos were randomly placed in positions 1 through 6 in both of the lineups. The lineups are included in Appendix A.

Before seeing the lineup, participants were told by the researcher, "Now you will be shown a lineup. Choose the person who you believe committed the burglary in the video that you saw previously. Indicate your choice using the numbers above or below that person's photo (1-6). If you do not believe that the person that you saw in the video is in the lineup, then you may choose, 7 (the person is not in the lineup)."

Demographic information. Participants were prompted to provide demographic information, including age, race, and gender; see Appendix A.

Procedure

After arriving at the laboratory, each participant read an information sheet and provided consent to participate in the study. Participants were then shown a 30 s video-clip depicting a female perpetrator burglarizing a home. Immediately following the video, participants engaged in one of the three recall instruction conditions for 2 min. Next, participants were shown either a target present or target absent lineup and were asked to either select a suspect from the lineup or indicate that the perpetrator was not present. Finally, participants completed a brief demographic questionnaire, were debriefed, and thanked for their participation.

Results

The results of Experiment 1 are reported in three main sections. In Section I, the effects of recall instructions on measures of identification accuracy are reported. In Section II, the effects of recall instructions on measures of description accuracy are reported. Finally, in Section III, the effects of recall instructions on lineup identification response times are reported. All analyses were performed using IBM SPSS Statistics Software.

I. Recall Instructions and Identification Accuracy

First, measures of discriminability and response criteria for each recall instruction condition are reported. Next, the effects of recall instructions on identification accuracy in target present (hits) and target absent (false alarms) lineups are reported. As previously stated, a hit is defined as a correct identification in a target present lineup, and a false alarm is defined as an incorrect identification in a target absent lineup. There are, of course, other types of identification errors that can be made in target present lineup identifications. While a hit is the only type of correct identification in a target present lineup, there are two possible identification errors: 1) an incorrect identification which is defined as an identification of a person, or foil, other than the correct target, and 2) a miss which is defined as failing to identify the correct target, or rejecting the lineup (‘the person is not in the lineup’). There has been a great deal of controversy regarding the possible significance of incorrect identification rates and miss rates (e.g., Kaye, 1986; Wixted and Mickes, 2012; Wells & Lindsay, 1980); however, consensus appears to place more importance on hit rates and false alarm rates (e.g., Clark, 2005; Palmer & Brewer, 2012). One reason for the emphasis on hit and false alarm rates is due to the low frequency of occurrence of incorrect identification and miss rates. Typically, incorrect identification rates and miss rates are so infrequent, that statistical analysis is not feasible. Another reason why recent research focus on hit and false rates is due to the types of analyses researchers are using to analyze identification accuracy. For instance, various researchers construct receiver operating curves (ROC analysis) to examine identification accuracy which depend upon the hit and false rates (e.g., Clark, 2003; Wixted & Mickes, 2014). Because the incorrect identification and miss rates were so low in the present studies (12.4% and 24.8% respectively), analyses were performed on hit and false alarm rates only.

Discriminability and response criteria. The adjusted false alarm rate was used in the computations of discriminability and response criteria. That is, the observed false alarm rate from the target absent lineup was divided by the total number of members in the lineup (Mickes, Flowe, & Wixted, 2012). In order to examine the influence of the recall instructions on participants’ discriminability of suspects in the lineup identification task, *d’* measures for each of the three recall instruction conditions were computed. Discriminability was highest in the ‘no recall’ instruction condition, *d’* = 2.75, intermediate in the ‘general recall’ instruction condition, *d’* = 2.19, and lowest in the ‘face recall’ instruction condition, *d’* = 0.87. See Table 1 for *d’* measures as a function of recall instruction condition.

In order to examine the influence of recall instructions on participants' response criteria in the lineup identification task, *C* measures for each of the three recall instruction conditions were computed. Response criterion was most conservative in the 'face recall' instruction condition, $C = 0.85$, intermediate in the 'general recall' instruction condition, $C = 0.66$, and was most liberal in the 'no recall' instruction condition, $C = 0.38$. The *C* measures as a function of recall instruction condition are reported in Table 1.

Identification accuracy. In order to examine the influence of recall instructions on identification accuracy, a logistic regression consisting of recall instructions (general recall, face recall, or no recall) was conducted separately for both *hits* and *false alarms*.

Hits. The percentage of hits as a function of recall condition are displayed in Table 1. There was a statistically significant effect of recall instructions on hits, Wald $\chi^2(2) = 16.30$, $p < .001$. The odds of making a hit were significantly greater for participants in the 'no recall' instruction condition than for participants in the 'face recall' instruction condition, Wald $\chi^2(1) = 15.52$, $p < .001$, OR = 9.86. Additionally, the odds of making a hit were significantly greater for participants in the 'general recall' instruction condition than for participants in the 'face recall' instruction condition, Wald $\chi^2(1) = 6.81$, $p = .009$, OR = 3.85. The odds of making a hit, while relatively large, were not statistically different between participants in the 'no recall' instruction condition and participants in the 'general recall' instruction condition, Wald $\chi^2(1) = 2.78$, $p = .095$, OR = 2.58. Thus, verbal overshadowing was observed when participants were asked to provide a specific description of the face, but not when participants were asked to provide a general description of the person's appearance.

False alarms. The percentage of false alarms as a function of recall condition are displayed in Table 1. The results demonstrated that, similar to the results for hits, recall instructions also had a statistically significant impact on participants' false alarm rates, Wald $\chi^2(2) = 12.67$, $p = .002$. The odds of making a false alarm were significantly greater for participants in the 'face recall' instruction condition than for participants in the 'no recall' instruction condition, Wald $\chi^2(1) = 9.31$, $p = .002$, OR = 4.85. Additionally, the odds of making a false alarm were significantly greater for participants in the 'face recall' instruction condition than for participants in the 'general recall' instruction condition, Wald $\chi^2(1) = 8.33$, $p = .004$, OR = 4.49. However, the odds making a false alarm were not different between participants in the 'no recall' instruction condition and participants in the 'general recall' instruction condition, Wald $\chi^2(1) = 0.02$, $p = .893$, OR = 1.10. Thus, as was observed for hits, participants' false alarm rates were affected by providing a specific description of the perpetrator's face, but not by providing a more general description of the suspect.

Recall instructions and identification accuracy results summary. Overall, the observed VOE found in previous studies was replicated in Experiment 1. Specifically, when participants were instructed to describe the facial features of a burglar rather than provide a general description of the burglar's appearance or not provide a description at all, then their identification accuracy was worse, their discriminability was lower, and their response criteria was more conservative.

II. Recall Instructions and Description Accuracy

First, the measures of description accuracy for the 'general recall' and 'face recall' instruction conditions are reported. Next, the overall relationship between description accuracy and identification accuracy are examined. Last, the relationship between description accuracy and identification accuracy separately for the 'general recall' and 'face recall' instruction conditions are reported.

Description accuracy. The internal consistencies of the four description accuracy measures were assessed. Internal consistency reflects the degree to which judgments about a description are consistent across observations or items thought to reflect the same descriptive dimension. Fifty (approximately 25%) of the descriptions were scored by the author and a research assistant. An inter-rater reliability analysis using the Kappa statistic was performed to determine consistency among the researchers. The interrater reliabilities for the four measures of description accuracy ranged from .77 to 1.00. According to Landis and Koch (1977), these Kappa values reflect a substantial measure of agreement between the two raters.

In order to examine the relationship between recall instructions and description accuracy, independent samples *t*-tests were performed separately for each of the four measures of description accuracy. The number of descriptors was significantly higher for participants in the 'general recall' instruction condition ($M = 8.80$, $SD = 2.63$) compared to participants in the 'face recall' instruction condition ($M = 7.35$, $SD = 2.91$), $t(138) = 3.08$, $p = .003$, $d = 0.523$. In addition, the number of correct descriptors was significantly higher for participants in the 'general recall' instruction condition ($M = 2.94$, $SD = 1.10$) compared to participants in the 'face recall' instruction condition ($M = 2.48$, $SD = 1.45$), $t(138) = 2.13$, $p = .035$, $d = 0.357$. Conversely, the number of incorrect descriptors was significantly higher for participants in the 'face recall' instruction condition ($M = 0.52$, $SD = 0.73$) compared to participants in the 'general recall' instruction condition ($M = 0.25$, $SD = 0.43$), $t(138) = 2.69$, $p = .008$, $d = -0.451$. Furthermore, the number of subjective descriptors was significantly higher for participants in the 'general recall' instruction condition ($M = 5.61$, $SD = 2.47$) compared to participants in the 'face recall' instruction condition ($M = 4.35$, $SD = 2.51$), $t(138) = 2.89$, $p = .003$, $d = 0.506$. Descriptive results for each recall instruction condition are displayed in Table 2.

Overall relationship between description accuracy and identification accuracy. The overall relationship between description accuracy and identification accuracy was examined. In order to assess the overall relationship between description accuracy and identification accuracy, the four measures of description accuracy were correlated to both hits and false alarms. None of the measures of description accuracy were related to hits or false alarms, *r*s ranged from -0.21 - 0.13, *p*s ranged from .089 - .846.

'General recall' instruction condition: relationship between description accuracy and identification accuracy. Next, the relationship between description accuracy and identification accuracy for the 'general recall' instruction condition was examined. None of the measures of description accuracy were related to hits or false alarms, *r*s ranged from -0.30 - 0.15, *p*s ranged from .077 - .900.

'Face recall' instruction condition: relationship between description accuracy and identification accuracy. The relationship between description accuracy and identification accuracy for the 'face recall' instruction condition was examined next. The number of descriptors and false alarms was positively correlated, $r(39) = 0.32$, $p = .045$. None of the other measures of description accuracy were related to hits or false alarms, *r*s ranged from -0.22 - 0.33, *p*s ranged from .066 - .929.

Recall instructions and description accuracy results summary. Results revealed that participants in the 'general recall' instruction condition had more correct descriptors, subjective descriptors, and number of descriptors compared to participants in the 'face recall' instruction condition. Participants in the 'face recall' instruction condition had more incorrect descriptors. In addition, recall instructions were found to be related to description accuracy. However, no evidence that description accuracy was related to identification accuracy was found despite the earlier results demonstrating a relationship between recall instructions and identification accuracy.

III. Recall Instructions and Lineup Identification Response Time

First, the relationship between recall instructions and lineup identification response time across lineups is reported: that is, the average amount of time it took participants in each recall condition to make an identification regardless of whether they were given a target present or target absent lineup. Next the overall relationship between identification response times and identification accuracy is reported. The relationship between identification response times and identification accuracy specifically for each of the three recall instruction conditions is also reported.

Lineup identification response time. The relationship between recall instructions and participants' lineup identification response times across both target present and target absent lineups was examined: that is, the amount of time in seconds it took participants to make their lineup identification decision despite lineup type. A single factor ANOVA consisting of recall instructions was conducted using identification response times as the dependent variable. Results revealed a significant effect of recall instructions on identification response times, $F(2, 209) = 15.08$, $p < .001$, $\eta_p^2 = 0.126$. Post-hoc analyses using LSD revealed that participants in the 'no recall' instruction condition had faster identification response times ($M = 10.07$, $SD = 7.15$) compared to participants in the 'general recall' instruction condition ($M = 15.35$, $SD = 10.41$), $p = .001$, and participants in the 'face recall' instruction condition ($M = 18.81$, $SD = 10.86$), $p < .001$. Participants in the 'general recall' instruction condition had faster identification response times compared to participants in the 'face recall' instruction condition, $p = .034$. Thus, overall, participants were faster to identify a perpetrator from a lineup when they had not previously attempted to provide a description of the person prior to the lineup. Response times for each recall instruction condition are reported in Table 1.

Overall relationship between lineup identification response time and identification accuracy. To assess the overall relationship between lineup identification response time and identification accuracy, lineup identification response time was correlated to hit and false alarm rates across all recall instruction conditions. Lineup identification response times were not related to hits or false alarms, *r*s ranged from -0.19 – 0.14, *p*s ranged from .058 - .160. See Table 3 for overall identification response time for hits and false alarms.

'No recall' instruction condition: relationship between lineup identification response time and identification accuracy. Next, whether there was a relationship between lineup identification response time and identification accuracy for the 'no recall' instruction condition was examined. Lineup identification response time was correlated with hits and false alarm rates in the 'no recall' instruction condition and revealed that response time was negatively correlated with hits, $r(37) = -0.53$, $p < .001$. In the 'no recall' instruction condition, participants who made correct identifications in the target present lineup had faster response times ($M = 8.22$, $SD = 6.07$) compared to participants who failed to identify the perpetrator ($M = 18.30$, $SD = 6.91$). Additionally, response time was positively correlated to false alarms, $r(35) = 0.34$, $p = .043$. Participants who made incorrect identifications in the target absent lineup had slower response times ($M = 14.77$, $SD = 5.69$) compared to participants who correctly rejected the lineup ($M = 8.95$, $SD = 7.16$). Identification response times for hits and false alarms in the 'no recall' instruction condition are displayed in Table 3.

'General recall' instruction condition: relationship between lineup identification response time and identification accuracy. Similar to the 'no recall' instruction condition, whether there was a relationship between lineup identification response time and identification accuracy for the ‘general recall’ instruction condition was examined. Lineup identification response time was correlated to hits and false alarms in the ‘general recall’ instruction condition. Response time was positively correlated to false alarms, $r(33) = 0.35$, $p = .044$. In the ‘general recall’ instruction condition, participants who made incorrect identifications in the target absent lineup had slower response times ($M = 20.05$, $SD = 11.60$) compared to participants who correctly rejected the lineup ($M = 13.39$, $SD = 6.30$). Response time was not related to hits. Identification response times for hits and false alarms in the ‘general recall’ instruction condition are displayed in Table 3.

‘Face recall’ instruction condition: relationship between lineup identification response time and identification accuracy. Last, whether there was a relationship between lineup identification response time and identification accuracy for the ‘face recall’ instruction condition was examined. Lineup identification response time was correlated to hits and false alarms in the ‘face recall’ instruction condition and revealed that response time was negatively correlated to false alarms, $r(39) = -0.52$, $p < .001$. In the ‘face recall’ instruction condition, participants who made incorrect identifications in the target absent lineup had faster response times ($M = 16.77$, $SD = 9.50$) compared to participants who correctly rejected the lineup ($M = 24.60$, $SD = 10.66$). Lineup identification response time was not related to hits. See Table 3 for identification response times for hit and false alarms in the ‘face recall’ instruction condition.

Lineup identification response time and identification accuracy results summary. When examining the relationship between lineup identification response time and identification accuracy regardless of lineup type, participants in the ‘no recall’ instruction condition had faster response times compared to the ‘general recall’ and ‘face recall’ instruction conditions. When considering lineup type, participants in the ‘no recall’ condition who correctly identified the perpetrator in the target present lineup had faster response times compared to participants who failed to identify the perpetrator. Furthermore, participants who incorrectly identified the perpetrator in the target absent lineup had slower response times compared to participants who correctly rejected the lineup. Participants in the ‘general recall’ instruction condition who incorrectly identified the perpetrator in the target absent lineup had slower response times compared to participants who correctly rejected the lineup. Participants in the ‘face recall’ instruction condition who incorrectly identified the perpetrator in the target absent lineup had faster response times compared to participants who correctly rejected the lineup.

Discussion

Identification Accuracy

The results of Experiment 1 revealed that participants in the ‘face recall’ instruction condition had poorer identification accuracy (hit rate = 34%, false alarm rate = 10%) compared to participants in the ‘no recall’ instruction condition (hit rate = 84%, false alarm rate = 4%). And while the ‘no recall’ instruction condition appeared to have better identification accuracy compared to the ‘general recall’ instruction condition (hit rate = 67%, false alarm rate = 4%), that difference in identification accuracy was not statistically significant.

This finding was also supported by the discriminability and response criteria results. Examining the extent to which verbal descriptions influenced discriminability and response criteria was another aim of Experiment 1. Verbal descriptions effected discriminability: participants in the ‘face recall’ instruction condition had the lowest measure of discriminability ($d' = 0.87$), followed by the ‘general recall’ instruction condition ($d' = 2.19$) and the ‘no recall’ instruction condition ($d' = 2.75$). Verbal descriptions were also related to response criteria:

participants in the ‘no recall’ instruction condition had the most liberal measure of response criteria (*C* = 0.38), followed by the ‘general recall’ instruction condition (*C* = 0.66) and then the ‘face recall’ instruction condition (*C* = 0.85).

Overall, results demonstrate a relationship between recall instructions and identification accuracy and are consistent with previous research examining the influence of certain types of recall instructions on identification accuracy (Clare & Lewandowsky, 2004; Finger & Pezdek, 1999; Meissner, 2002; Meissner et al., 2001; Sauerland, Holub, & Sporer, 2008; Smith & Flowe, 2015). For example, Smith and Flowe (2015) provided participants with one of four types of recall instructions (no description, standard description, warning description, or forced description). Smith and Flowe’s ‘standard recall’ condition was very similar to the ‘face recall’ instruction condition employed in the current study (see Table 4). The researchers found an influence of recall instructions on identification accuracy. Participants in the ‘standard recall’ condition had poorer identification accuracy compared to participants in the ‘no recall’ condition. Because Smith and Flowe reported the hit and false alarms rates for each recall instruction condition, it was possible to compute *d’* and *C* and compare their discriminability and response criteria results to the current study’s. Smith and Flowe’s measures of discriminability and response criteria in the ‘no recall’ condition (*d’* = 1.61, *C* = 0.48) indicated that participants had a higher measure of discriminability and more liberal measure of response criteria compared to participants in the ‘standard recall’ condition (*d’* = 1.33, *C* = 0.51). Consistent with Smith and Flowe, the current study’s measures of discriminability and response criteria in the ‘no recall’ instruction condition (*d’* = 2.75, *C* = 0.38) also indicated that participants had a higher measure of discriminability and more liberal measure of response criterion compared to participants in the ‘face recall’ instruction condition (*d’* = 0.87, *C* = 0.85).

Description Accuracy

The observation that description accuracy was not related to identification accuracy is potentially important regarding theory development for the VOE. Results revealed that recall instructions were related to both description accuracy and identification accuracy. Regarding description accuracy, results also revealed that participants in the ‘general recall’ instruction condition had more correct descriptors, subjective descriptors, and number of descriptors compared to the ‘face recall’ condition. Additionally, participants in the ‘face recall’ condition had more incorrect descriptors compared to the ‘general recall’ condition. And as previously discussed, regarding identification accuracy, participants in the ‘general recall’ condition had better identification accuracy compared to participants in the ‘face recall’ condition.

While the expected relationships between recall instructions and both description and identification accuracy (hits and false alarms) were found, it was surprising to find that description accuracy was not related to identification accuracy. The only relationship found was between number of descriptors (the number of adjectives used in their descriptions) and false alarms. When looking at the ‘face recall’ condition, the number of descriptors was negatively associated with false alarms. As number of descriptors is not a measure of accuracy, why it would be related to identification accuracy is not clear.

In accord with previous attempts to assess which aspects of verbal descriptions may be related to identification accuracy, a number of techniques to measure description accuracy (i.e., correct descriptors, incorrect descriptors, and subjective descriptors) were used in the present study. Experiment 1 failed to establish a meaningful relationship between description accuracy and identification accuracy. The results are inconsistent with the previous researchers that have found small but reliable correlations between similar measures of description accuracy and identification accuracy (Meissner, Sporer, Schooler, 2006; Sporer, 1996). Experiment 1's results provide support for the hypothesis that both describing and identifying a face may rely on different underlying mental representation. This support is consistent with the TIPS account for the VOE and inconsistent with the RBI account.

Lineup Identification Response Time

The identification response time data in Experiment 1 could be potentially important for theory development. Results of Experiment 1 provide evidence that supports a relationship between identification response time and identification accuracy. More specifically, there was no overall relationship between identification accuracy and identification response time (Table 3). However, when the relationship between identification accuracy and identification response time for each recall instruction condition was examined, a potentially interesting effect was found. In the 'no recall' condition, participants who correctly identified the perpetrator in a target present lineup had faster response times compared to participants who did not; participants who made an incorrect identification in a target absent lineup had slower response times compared to participants who correctly rejected the lineup (Table 3). The results are consistent with previous studies that have also found faster response times associated with correct identifications and slower response times associated with incorrect identifications (Brewer et al., 2006; Bruer & Pozzulo, 2014; Dunning & Perretta, 2002; Sauerland & Sporer, 2009; Sporer, 1993; Weber et al., 2004).

However, when looking at the 'general recall' condition, a relationship between hits and response time was not observed (Table 3). In other words, participants in the 'general recall' condition were no faster at correctly recognizing and identifying the target than when making an incorrect identification. It is not clear what to make of this finding. It is possible that the leveling off in response time may indicate that participants who were asked to recall the target's description were less sure about their identification decision. The same lack of relationship between response time and hits was also observed for the 'face recall' condition (Table 3). But interestingly, in the 'face recall' condition there was a relationship between response time and false alarms. Participants in the 'face recall' condition who made an incorrect identification in a target absent lineup had faster response times compared to participants who correctly rejected the lineup. Unlike the 'no recall' and 'general recall' conditions, which had faster responses when making correct identifications from target present lineups, the 'face recall' condition had faster response times when making incorrect identifications from target absent lineups. Why the 'face recall' conditions' response times were faster when making incorrect identifications from target absent lineups is difficult to explain. These results contradict previous research that suggests that witnesses who incorrectly identify a target from a target absent lineup are slower to reach a decision compared to witnesses who correctly reject a target absent lineup (Smith, Lindsay, & Pryke, 2000; Sporer, 1992, 1993). Furthermore, research by Weber et al. (2004) suggests that it can take anywhere from 5-29s for people to make a discrimination judgment amongst faces. Participants in the 'face recall' condition who incorrectly identified a perpetrator in a target absent lineup had an average response time of 16.77s which is consistent with Weber et al.'s finding. However, from our measure of discriminability for the 'face recall' condition (d' = 0.87), it was apparent that these participants were unable to discriminate amongst faces in either lineup.

Recall Instructions

It is apparent that the manipulation of the recall instructions had an influence on identification accuracy. In particular, a large difference in identification accuracy between the ‘general recall’ instruction condition and the ‘face recall’ instruction condition was observed. There is possible explanation for these observed differences: recall instruction length. Participants in the ‘general recall’ instruction condition were given brief instructions (9 words long) to describe the person from the video footage while participants in the ‘face recall’ instruction condition were given much longer instructions (73 words long). It is possible that the length of recall instructions influenced participants’ memories. Prior research suggests that people have a limited amount of information that they can hold in memory for a certain amount of time (e.g., Cowen, 2001; Gobet & Clarkson, 2004). It is possible that the longer length of the recall instructions that participants in the ‘face recall’ instruction condition had to remember was cognitively demanding and interfered with their original memory of the target person. In Table 3, the length of Experiment 1’s ‘face recall’ instruction condition is compared to previous verbal overshadowing studies similar recall instruction conditions. The length of those studies’ recall conditions (60, 88, and 72 words) were comparable to the length of the ‘face recall’ instruction condition (73 words). This may provide some insight as to why the ‘face recall’ instruction condition resulted in verbal overshadowing while the ‘general recall’ instruction condition did not in the present study. Recall instruction length was examined further in Experiment 2.

CHAPTER III: EXPERIMENT 2

In Experiment 2, both recall instruction type and recall instruction length was manipulated to determine if the VOE was a result the type of recall participants were asked to engage in, general versus face, or if it was a result of the nature and length of the recall instructions, short versus long. If the large VOE observed in Experiment 1 was due to the type of recall participants engaged in, that is, if recalling the target's facial features resulted in the VOE rather than just recalling the target's general appearance, then participants in the 'face recall' conditions should demonstrate the VOE. On the other hand, if the VOE in Experiment 1 was due to the length of the recall instructions, that is, if being given long, cued recall instructions resulted in the VOE rather than being given short recall instructions, then participants in the 'long recall' conditions should demonstrate the VOE.

As in Experiment 1, participants in Experiment 2 watched the burglar video, recalled the physical description of the burglar, and attempted to make an identification of the burglar from a lineup. Recall instruction type and recall instruction length were varied between participants. Participants in a 'general short recall' instruction condition provided a brief description of the perpetrator's physical appearance; participants in the 'general long recall' instruction condition provided a detailed description of the perpetrator's physical appearance; participants in the 'face short recall' instruction condition provided a brief description of the perpetrator's face; participants in the 'face long recall' instruction condition provided a detailed description of the perpetrator's facial features; and participants in a 'no recall' instruction condition did not describe the perpetrator.

Method

Design

The study was a 2 (recall instruction type: face, general) × 2 (recall instruction length: short, long) × 2 (lineup type: target present, target absent) between-groups factorial design with an outside 'no recall' instruction control condition. All participants were randomly assigned to conditions. The study was approved by the University of Mississippi's IRB.

Participants

The participants included a sample of 360 undergraduate students (66.90% female, overall M_{age} = 18.93). All participants received partial course credit for participating. This study was also administered using Qualtrics online surveying software.

Materials and Procedure

The materials and procedure used in Experiment 2 were the same as those used in Experiment 1, apart from the recall instructions that participants were given. As in Experiment 1, some participants were either told, "Please describe the person you saw in the video;" those participants were in the 'general short recall' instruction condition. Also, as in Experiment 1, some participants were told, "Please describe the face you saw in the video. Your task is to describe the person in such a way that your description would aid someone else in attempting to identify the person. Your description should focus on facial features. Write about the shape and size of the eyes, eyebrows, nose, ears, mouth, chin, etc. Try not to leave out any details about the face even if you think they are not important;" those participants were in the 'face long recall' instruction condition. In addition to these recall instructions, participants in the 'face short recall' instruction condition were told, "Please describe the face you saw in the video." Lastly, participants in the 'general long recall' instruction condition were told, "Please describe the person you saw in the video. Your task is to describe the person in such a way that your description would aid someone else in attempting to identify the person. Your description should focus on the person's appearance. Write about the person's height, weight, race, gender, hair color, hairstyle, clothing, etc. Try not to leave out any details about the person's appearance even if you think they are not important." Participants were provided 2 min to complete their descriptions. Participants in the 'no recall' instruction condition played a game of Tetris.

Results

As in Experiment 1, the results for Experiment 2 are also reported in three main sections. In Section I, results regarding the effects of recall instructions on identification accuracy are reported. In Section II, results regarding the effects of recall instructions on description accuracy are reported. Finally, in Section III, the effects of recall instructions on lineup identification response times are reported.

I. Recall Instructions and Identification Accuracy

First, measures of discriminability and response criteria for each recall instruction condition are examined. Next, the effects of recall instructions on identification accuracy in target present (hits) and target absent (false alarms) lineups are reported. As in Experiment 1, analyses on incorrect identifications and misses were not performed in Experiment 2 either, as those types were so infrequent, 13.3% and 31.1% respectively.

Discriminability and response criteria. To examine the influence of the recall instructions on participants' discriminability of suspects in the lineup identification task, *d'* measures for each of the five recall instruction conditions were computed. Discriminability was highest in the 'no recall' instruction condition, *d'* = 2.49, followed by the 'general short recall,' *d'* = 1.76, 'general long recall,' *d'* = 1.45, 'face short recall,' *d'* = 1.15, and 'face long recall,' *d'* = 1.08, instruction conditions; see Table 5.

To examine the influence of recall instructions on participants' response criteria in the lineup identification task, *C* measures for each of the five recall instruction conditions were computed. Response criterion was most conservative in the 'general long recall' instruction condition, *C* = 0.75, followed by the 'face long recall,' *C* = 0.69, 'general short recall,' *C* = 0.68, 'face short recall,' *C* = 0.65, and 'no recall,' *C* = 0.41, instruction conditions; see Table 5.

Identification accuracy. To examine the influence of recall instructions on identification accuracy, a logistic regression consisting of recall instruction was conducted separately for both hit and false alarm rates.

Hits. The percentage of hits as a function of recall condition are displayed in Table 5. There was a significant effect of recall instructions on hits, Wald $\chi^2(4) = 11.02$, $p = .026$. The odds of making a hit were significantly greater for participants in the 'no recall' instruction condition than for participants in the other recall instruction conditions; the 'face long recall' condition, Wald $\chi^2(1) = 8.90$, $p = .003$, OR = 5.00, the 'face short recall' condition, Wald $\chi^2(1) = 7.73$, $p = .005$, OR = 4.47, the 'general long recall' condition, Wald $\chi^2(1) = 8.90$, $p = .008$, OR = 4.24, and the 'general short recall' condition, Wald $\chi^2(1) = 3.98$, $p = .046$, OR = 2.91. However, the odds of making a hit were not statistically different between participants in the 'general short recall,' 'general long recall,' 'face short recall,' and 'face long recall' conditions, Wald $\chi^2(1)$s ranged between 0.84-1.33, *p*s ranged between 0.249-0.909, ORs ranged between 0.58-1.54. Thus, in Experiment 2, verbal overshadowing was observed when participants were asked to provide a description of the suspect regardless of recall instruction type (general versus face) or recall instruction length (short versus long).

False alarms. The percentage of false alarms as a function of recall condition are displayed in Table 5. The results demonstrated that, similar to the results for hits, the recall instructions also had a statistically significant impact on participants' false alarm rates, Wald $\chi^2(4) = 14.26$, $p = .007$. Compared to participants in the 'no recall' condition, the odds of making a false alarm were significantly greater for participants in the 'face long recall' condition, Wald $\chi^2(1) = 8.20$, $p = .004$, OR = 4.17 and the 'face short recall' condition, Wald $\chi^2(1) = 7.50$, $p = .006$, OR = 3.85. Additionally, compared to participants in the 'general short recall' condition, the odds of making a false alarm were significantly greater for participants in the 'face long recall' condition, Wald $\chi^2(1) = 6.01$, $p = .014$, OR = 3.39 and the 'face short recall' condition, Wald $\chi^2(1) = 5.39$, $p = .020$, OR = 3.12. Also, compared to participants in the 'general long recall' condition, the odds of making a false alarm were significantly greater for participants in the 'face long recall' condition, Wald $\chi^2(1) = 3.98$, $p = .046$, OR = 2.67. However, the odds of making a false alarm were not statistically different for all other comparisons made between recall instructions conditions, Wald $\chi^2(1)$s ranged between 0.18-3.98, *p*s ranged between 0.063-0.675, ORs ranged between 0.64-3.39. Overall, participants' false alarm rates were affected by providing a specific description of the perpetrator's face, but not by providing a more general description of the suspect depicted in the video.

Recall instructions and identification accuracy results summary. As in Experiment 1, the verbal overshadowing effect was also found in Experiment 2. Regardless of recall instruction type (general verses face) and recall instruction length (short versus long), identification accuracy in target present lineups was poorer in all 4 recall instructions conditions compared to the 'no recall' control condition. Identification accuracy was poorest in the 'face long recall' instruction condition, followed by the 'face short recall,' 'general long recall,' and 'general short recall' instruction conditions. However, regarding identification accuracy from target absent lineups, participants in the 'face long recall' and 'face short recall' instruction conditions had poorer identification accuracy compared to the 'no recall,' 'general short recall,' and 'general long recall' instruction conditions.

II. Recall Instructions and Description Accuracy

First, the relationship between recall instructions and description accuracy are reported. Next, the overall relationship between description accuracy and identification accuracy are reported. Reported last is the relationship between description accuracy and identification accuracy separately for the 'general short,' 'general long,' 'face short,' and 'face long' recall instruction conditions.

Description accuracy. As in Experiment 1, the internal consistency of the four description accuracy measures was assessed. One hundred-thirty (approximately 45%) of the descriptions were scored by the author and a research assistant. The interrater reliabilities for the 4 measures of description accuracy ranged from .77 to 0.97 demonstrating a considerable measure of agreement between the two raters (Landis & Koch, 1977).

To examine the relationship between recall instructions and description accuracy, factorial ANOVAs consisting of recall type (general versus face) and recall length (short versus long) were performed separately for each of the 4 measures of description accuracy. Regarding the number of descriptors, there were main effects of recall type and recall length, however, there was no interaction. A main effect of recall type was found, $F(1, 284) = 18.33$, $p < .001$, $\eta_p^2 = 0.061$, such that participants in the 'general recall' conditions had a higher number of descriptors ($M = 7.90$. $SD = 2.78$) than participants in the 'face recall' conditions ($M = 6.65$, $SD = 2.95$). A main effect of recall length was found, $F(1, 284) = 84.73$, $p < .001$, $\eta_p^2 = 0.230$, such that participants in the ‘long recall’ conditions had a higher number of descriptors (*M* = 8.65. *SD* = 2.68) than participants in the ‘short recall’ conditions (*M* = 5.92, *SD* = 2.52). Regarding correct descriptors, there were main effects of recall type and recall length, however, there was no interaction. A main effect of recall type was found, $F(1, 284) = 108.81$, $p < .001$, $\eta_p^2 = 0.277$, such that participants in the ‘general recall’ conditions had more correct descriptors (*M* = 5.98. *SD* = 2.34) than participants in the ‘face recall’ conditions (*M* = 3.63, *SD* = 1.75). A main effect of recall length was found, $F(1, 284) = 45.60$, $p < .001$, $\eta_p^2 = 0.138$, such that participants in the ‘long recall’ conditions had more correct descriptors (*M* = 5.56. *SD* = 2.26) than participants in the ‘short recall’ conditions (*M* = 4.05, *SD* = 2.25). Regarding incorrect descriptors, there were main effects of recall type and recall length, however, there was no interaction. A main effect of recall type was found, $F(1, 284) = 7.09$, $p = .009$, $\eta_p^2 = 0.024$, such that participants in the ‘general recall’ conditions had more incorrect descriptors (*M* = 0.66. *SD* = 0.78) than participants in the ‘face recall’ conditions (*M* = 0.45, *SD* = 0.60). A main effect of recall length was found, $F(1, 284) = 21.99$, $p < .001$, $\eta_p^2 = 0.072$, such that participants in the ‘long recall’ conditions had more incorrect descriptors (*M* = 0.74. *SD* = 0.75) than participants in the ‘short recall’ conditions (*M* = 0.37, *SD* = 0.60). Regarding subjective descriptors, there were main effects of recall type and recall length, however, there was no interaction. A main effect of recall type was found, $F(1, 284) = 46.25$, $p < .001$, $\eta_p^2 = 0.140$, such that participants in the ‘face recall’ conditions had more subjective descriptors (*M* = 2.57. *SD* = 1.91) than participants in the ‘general recall’ conditions (*M* = 1.27, *SD* = 1.37). A main effect of recall length was found, $F(1, 284) = 19.55$, $p < .001$, $\eta_p^2 = 0.064$, such that participants in the ‘long recall’ conditions had more subjective descriptors (*M* = 2.35. *SD* = 2.00) than participants in the ‘short recall’ conditions (*M* = 1.50, *SD* = 1.43). Descriptive results for each recall instruction condition are displayed in Table 6.

Overall relationship between description accuracy and identification accuracy. The overall relationship between description accuracy and identification accuracy was examined. To assess the overall relationship between description accuracy and identification accuracy, the four measures of description accuracy were correlated to hits and false alarms across all recall instruction conditions. Subjective descriptors and hits was negatively correlated, $r(145) = -0.19$, $p = .026$. Number of descriptors and false alarms was negatively correlated, $r(145) = -0.18$, $p = .030$. Additionally, correct descriptors and false alarms was negatively correlated, $r(145) = -0.18$, $p = .029$. None of the other measures of description accuracy were related to hits or false alarms, *r*s ranged from -0.12 – 0.03, *p*s ranged from 0.16 – 0.93.

'General short recall' instruction condition: relationship between description accuracy and identification accuracy. Next, whether there was a relationship between description accuracy and identification accuracy in the 'general short recall' instruction condition was examined. Number of descriptors and false alarms was negatively correlated, $r(38) = -0.42$, $p = .013$. Additionally, correct descriptors and false alarms was negatively correlated, $r(38) = -0.43$, $p = .009$. None of the other measures of description accuracy were related to hits or false alarms, *r*s ranged from -0.09 – 0.24, *p*s ranged from 0.14 – 0.91.

'General long recall' instruction condition: relationship between description accuracy and identification accuracy. Next, whether there was a relationship between description accuracy and identification accuracy in the 'general long recall' instruction condition was examined. None of the measures of description accuracy were related to hits or false alarms, *r*s ranged from -0.03 – 0.28, *p*s ranged from 0.11 – 0.86.

'Face short recall' instruction condition: relationship between description accuracy and identification accuracy. Next, whether there was a relationship between description accuracy and identification accuracy in the 'face short recall' instruction condition was examined. None of the measures of description accuracy were related to hits or false alarms, *r*s ranged from -0.32 – 0.15, *p*s ranged from 0.05 – 0.82.

'Face long recall' instruction condition: relationship between description accuracy and identification accuracy. Next, whether there was a relationship between description accuracy and identification accuracy in the 'face long recall' instruction condition was examined. Number of descriptors and hits was negatively correlated, $r(36) = -0.33$, $p = 0.47$. None of the other measures of description accuracy were related to hits or false alarms, *r*s ranged from -0.32 – -0.10, *p*s ranged from 0.06 – 0.56.

Recall instructions and description accuracy results summary. Results revealed that both recall type and recall length were related to description accuracy. Regarding recall type, participants in the 'general recall' conditions had more correct descriptors, incorrect descriptors, and a higher number of descriptors compared to participants in the face recall conditions. Participants in the 'face recall' conditions had more subjective descriptors. Regarding recall length, participants in the long recall conditions had more correct, incorrect, subjective, and a higher number of descriptors compared to participants in the short recall conditions. In addition to the observed relationship between recall instructions and identification accuracy, some evidence for a relationship between description accuracy and identification accuracy was also found. Specifically, when examining the relationship between description accuracy and identification accuracy overall (across all recall instruction conditions), more subjective descriptors were associated with fewer hits. Additionally, more correct descriptors were associated with fewer false alarms, and a higher number of descriptors were associated with fewer false alarms. Regarding the 'general short' recall instruction condition, more correct descriptors as well as a higher number of descriptors were also associated with fewer false alarms. The only other association observed was for the 'face long recall' instruction condition; a higher number of descriptors were associated with fewer hits.

III. Recall Instructions and Lineup Identification Response Time

First, the differences in identification response time as a function of recall instruction condition are reported. Next, the overall relationship between identification response times and identification accuracy across all five recall instruction conditions are reported. Last, the relationship between identification response times and identification accuracy specifically for each of the five recall instruction conditions are reported.

Lineup identification response time. To examine the influence of recall instructions on identification response time, a single factor ANOVA consisting of all five recall instructions was conducted using identification response times (in seconds) as the dependent variable. The analysis was conducted examining response time across both target present and target absent lineups. Results revealed a significant effect of recall instructions on identification response times, $F(4, 355) = 22.50$, $p < .001$, $\eta_p^2 = 0.202$. Post-hoc analyses using LSD revealed that participants in the 'no recall' instruction condition had faster identification response times (*M* = 8.83, *SD* = 3.32) compared to participants in the 'general short recall' (*M* = 15.36, *SD* = 7.00), 'general long recall' (*M* = 16.86, *SD* = 5.26), 'face short recall' (*M* = 16.74, *SD* = 6.85), and 'face long recall' (*M* = 15.71, *SD* = 6.67) instruction conditions. Identification response times between the 'general short recall,' 'general long recall,' 'face short recall,' and 'face long recall' instruction conditions were not different, *p*'s ranged from 0.14 – 0.93. Thus, overall, participants were fastest to identify a perpetrator from a lineup when they had not previously attempted to recall information about the person prior to the lineup. Response times for each recall instruction condition are reported in Table 5.

Overall relationship between lineup identification response time and identification accuracy. First, whether there was a relationship between lineup identification response time and identification accuracy across all recall instruction conditions was examined. Lineup identification response time was correlated to hit and false alarm rates across all recall instruction conditions and revealed that response time was negatively correlated to hits, $r(180) = -0.43$, $p < .001$. Across all recall instruction conditions, participants who made correct identifications in the target present lineup had faster response times ($M = 11.70$, $SD = 5.38$) compared to participants who failed to identify the perpetrator ($M = 17.37$, $SD = 6.59$). Response time was not correlated to false alarms, $r(180) = 0.11$, $p = .149$. Participants who made incorrect identifications in the target absent lineup had response times ($M = 15.94$, $SD = 6.14$) similar to participants who correctly rejected the lineup ($M = 14.49$, $SD = 7.29$). Identification response times for hits and false alarms across all recall instruction conditions are displayed in Table 7.

'No recall' instruction condition: relationship between lineup identification response time and identification accuracy. Next, whether there was a relationship between lineup identification response time and identification accuracy for the 'no recall' instruction condition was examined. Lineup identification response time was correlated to hit and false alarm rates in the 'no recall' instruction condition and revealed that response time was negatively correlated to hits, $r(35) = -0.60$, $p < .001$. In the 'no recall' condition, participants who made correct identifications in the target present lineup had faster response times ($M = 7.03$, $SD = 2.29$) compared to participants who failed to identify the perpetrator ($M = 12.05$, $SD = 4.37$). Response time was not correlated to false alarms, $r(37) = 0.15$, $p = .377$. Participants who made incorrect identifications in the target absent lineup had response times (*M* = 10.22, *SD* = 3.79) similar to participants who correctly rejected the lineup (*M* = 9.28, *SD* = 2.69). Identification response times for hits and false alarms in the 'no recall' instruction condition are displayed in Table 7.

'General short recall' instruction condition: relationship between lineup identification response time and identification accuracy. Next, whether there was a relationship between lineup identification response time and identification accuracy for the 'general short recall' instruction condition was examined. Lineup identification response time was correlated to hit and false alarm rates in the 'general short recall' instruction condition. Response time was not correlated to hits, $r(38) = -0.31$, $p = .059$, or false alarms, $r(37) = 0.06$, $p = .730$. In the 'general short recall' condition, participants who made correct identifications in the target present lineup had response times (*M* = 12.76, *SD* = 5.68) similar to participants who failed to identify the perpetrator (*M* = 18.08, *SD* = 10.94). Additionally, participants who made incorrect identifications in the target absent lineup had response times (*M* = 16.13, *SD* = 4.37) similar to participants who correctly rejected the lineup (*M* = 15.54, *SD* = 5.12). Identification response times for hits and false alarms in the 'general short recall' instruction condition are displayed in Table 7.

'General long recall' instruction condition: relationship between lineup identification response time and identification accuracy. Next, whether there was a relationship between lineup identification response time and identification accuracy for the 'general long recall' instruction condition was examined. Lineup identification response time was correlated to hit and false alarm rates in the 'general long recall' instruction condition. Response time was not correlated to hits, $r(35) = -0.32$, $p = .065$, or false alarms, $r(35) = 0.13$, $p = .464$. In the 'general long recall' condition, participants who made correct identifications in the target present lineup had similar response times (*M* = 14.48, *SD* = 5.18) to participants who failed to identify the perpetrator (*M* = 17.70, *SD* = 5.05). Additionally, participants who made incorrect identifications in the target absent lineup had response times (*M* = 18.37, *SD* = 5.31) similar to participants who correctly rejected the lineup (*M* = 17.01, *SD* = 5.72). Identification response times for hits and false alarms in the 'general long recall' instruction condition are displayed in Table 7.

'Face short recall' instruction condition: relationship between lineup identification response time and identification accuracy. Next, whether there was a relationship between lineup identification response time and identification accuracy for the 'face short recall' instruction condition was examined. Lineup identification response time was correlated to hit and false alarm rates in the 'face short recall' instruction condition and revealed that response time was negatively correlated to hits, $r(36) = -0.40$, $p = .016$. In the 'face short recall' condition, participants who made correct identifications in the target present lineup had faster response times (*M* = 14.03, *SD* = 5.17) compared to participants who failed to identify the perpetrator (*M* = 18.20, *SD* = 4.71). Response time was not correlated to false alarms, $r(37) = -0.05$, $p = .757$. Participants who made incorrect identifications in the target absent lineup had response times (*M* = 16.93, *SD* = 6.79) similar to participants who correctly rejected the lineup (*M* = 17.81, *SD* = 10.46). Identification response times for hits and false alarms in the 'face short recall' instruction condition are displayed in Table 7.

'Face long recall' instruction condition: relationship between lineup identification response time and identification accuracy. Next, whether there was a relationship between lineup identification response time and identification accuracy for the 'face long recall' instruction condition was examined. Lineup identification response time was correlated to hit and false alarm rates in the ‘face long recall’ instruction condition and revealed that response time was negatively correlated to hits, $r(36) = -0.43$, $p = .008$. In the ‘face long recall’ condition, participants who made correct identifications in the target present lineup had faster response times ($M = 12.95$, $SD = 4.82$) compared to participants who failed to identify the perpetrator ($M = 17.57$, $SD = 4.97$). Response time was not correlated to false alarms, $r(36) = 0.03$, $p = .848$. Participants who made incorrect identifications in the target absent lineup had response times ($M = 16.20$, $SD = 6.62$) similar to participants who correctly rejected the lineup ($M = 15.66$, $SD = 10.17$). Identification response times for hits and false alarms in the ‘face long recall’ instruction condition are displayed in Table 7.

Lineup identification response time and identification accuracy results summary. When examining the influence of recall instructions on identification response time regardless of lineup type, participants in the ‘no recall’ instruction condition had faster response times compared to all other recall instruction conditions. When examining the relationship between identification response time and identification accuracy across all lineup instruction conditions, participants who made correct identifications in target present lineups had faster response times than participants who failed to correctly identify the perpetrator. Additionally, participants in the ‘no recall,’ ‘face short recall,’ and ‘face long recall’ instruction conditions who made correct identifications in target present lineups had faster response times than participants who failed to correctly identify the perpetrator; however, those same results were not found in the ‘general short recall’ and ‘general long recall’ instruction conditions.

Discussion

Identification Accuracy

In Experiment 2, I investigated the influence of recall instructions on the VOE by manipulating both recall instruction type (general versus face) and recall instruction length (short versus long). Regarding hit rates, participants in all recall instruction conditions demonstrated verbal overshadowing compared to the ‘no recall’ instruction condition (hit rate = 80%): the ‘face long recall’ instruction condition had the lowest (hit rate = 44%), followed by the ‘face short recall’ (hit rate = 48%), ‘general long recall’ (hit rate = 49%), and ‘general short recall’ (hit rate = 58%) instruction conditions. Regarding false alarm rates, only participants in the ‘face recall’ instruction conditions demonstrated verbal overshadowing compared to the ‘no recall’ instruction condition (false alarm rate = 32%): the ‘face long recall’ instruction condition had the highest (false alarm rate = 67%), followed by the ‘face short recall’ instruction condition (false alarm rate = 65%); the false alarm rates for the ‘general short recall’ (false alarm rate = 38%) and ‘general long recall’ instruction conditions (false alarm rate = 43%) were not higher than the ‘no recall’ instruction condition. The results of the hit rate data suggest that recall of any kind resulted in verbal overshadowing: specifically, regardless of recall instruction type or recall instruction length, participants who provided a description of the target demonstrated the VOE when shown a target present lineup. However, the results of the false alarm data suggest that recall of the target’s facial features resulted in verbal overshadowing: specifically, regardless of recall instruction length, participants who provided a description of the target’s facial features demonstrated the VOE when shown a target absent lineup. The measures of discriminability and response criteria also support the hit and false alarm results. As displayed in Table 5, discriminability was lowest and response criteria was more conservative in the ‘face recall’ and ‘general recall’ instruction conditions compared to the ‘no recall’ instruction condition.

Description Accuracy

Description accuracy as well as the relationship between description accuracy and identification accuracy was assessed. Recall instructions were related to description accuracy, see Table 7. Regarding description accuracy for the recall instruction type factor, results revealed that participants in the ‘general recall’ instruction conditions had more correct descriptors, incorrect descriptors, and number of descriptors compared to the ‘face recall’ instruction conditions. However, participants in the ‘face recall’ instruction conditions had more subjective descriptors compared to the ‘general recall’ instruction condition. Regarding description accuracy for the recall instruction length factor, results revealed that participants in the ‘long recall’ instruction conditions had more correct descriptors, incorrect descriptors, subjective descriptors, and number of descriptors compared to the ‘short recall’ instruction conditions. This finding that the ‘long recall’ instruction conditions resulted in more of all description accuracy measures is not surprising as those participants were encouraged to produce longer descriptions, thus increasing their probability of producing more types of descriptors altogether (i.e., correct, incorrect, and subjective).

Regarding the relationship between description accuracy and identification accuracy, some potentially interesting results were found. First, when looking across recall instruction conditions a negative correlation between correct descriptors and false alarms was found demonstrating that more correct descriptors were associated with fewer false alarms. This same relationship between correct descriptors and false alarms was also observed in the ‘general short recall’ instruction condition. A second potentially interesting finding were the negative correlations between number of descriptors and false alarms observed over all recall instructions conditions and in the ‘face long recall’ instruction condition. The finding demonstrated that a higher number of descriptors was associated with fewer false alarms. Why number of descriptors would be related to identification accuracy is unclear, and how it relates to the VOE is also not entirely clear-cut; however, this decrease in false alarms might be explained by the tendency for participants in the ‘face long recall’ instruction condition to respond more conservatively during the lineup identification task.

Overall, a relationship between recall instruction conditions and both description accuracy and identification accuracy were found. However, when looking at the relationship between description accuracy and identification accuracy only certain associations were found; specifically, more correct descriptors were associated with fewer false alarms. And, as in Experiment 1, relationships between number of descriptors and identification accuracy were found, however, since number of descriptors is not a measure of accuracy per se, it is difficult to see how that relationship is diagnostic regarding how descriptions might relate to the VOE. Overall, it is not entirely clear how the results of Experiment 2 contribute to a better understanding of how the relationship between description accuracy and identification accuracy can explain the VOE.

Lineup Identification Response Time

The results of Experiment 2 provide evidence that may support a meaningful relationship between identification response time and identification accuracy. There was an overall relationship between identification response time and identification accuracy. Across all recall instruction conditions, participants who correctly identified the perpetrator in a target present lineup had faster response times compared to participants who did not. This same pattern was found when examining the relationship between identification response time and identification accuracy in the ‘no recall,’ ‘face short recall,’ and ‘face long recall’ instruction conditions. These results are consistent with that of Experiment 1 as well as previous studies that have also observed faster response times corresponding with correct identifications (Brewer et al., 2006; Bruer & Pozzulo, 2014; Dunning & Perretta, 2002; Sauerland & Sporer, 2009; Sporer, 1993; Weber et al., 2004).

While the relationship between response time and identification accuracy was observed in the ‘no recall’ and ‘face recall’ instruction conditions, the same relationship was not found for the ‘general recall’ instruction conditions. This lack of finding is also consistent with what was observed in Experiment 1. At this point, it is unclear why this relationship between response time and identification accuracy was not observed in the ‘general recall’ instruction conditions.

Recall Instructions

The manipulation of recall instructions in Experiment 2 had an influence on identification accuracy. Taking the identification accuracy results as a whole (hits, false alarms, *d’* and *C*), it appears that participants in both the ‘face recall’ and ‘general recall’ instruction conditions demonstrated verbal overshadowing compared to participants in the ‘no recall’ instruction condition. Regardless of recall instruction type and recall instruction length, participants who were asked to provide a description of the perpetrator had poorer identification accuracy, were less able to discriminate between the faces in the lineup and were more conservative when making a lineup identification; see Table 5.

The goal of Experiment 2 was to examine whether recall instruction length was responsible for the observed VOE found in Experiment 1 as well as those effects reported in previous research. Results of Experiment 2 suggest that recall instruction length had little effect on the VOE; thus, it appears that any cognitive demand or difficulty posed on participants by the word length of the recall instructions was not a reason for observed verbal overshadowing in Experiment 1's 'face recall' instruction condition. It may seem intuitive that longer and more specific recall instructions that require participants to think remember information for a certain amount of time may be a reason for poor subsequent recognition memory performance. This possibility is also consistent with previous research regarding the limited amount of information people can maintain (e.g., Gobet & Clarkson, 2004). However, results of Experiment 2 do not support this notion; rather, the results of Experiment 2 suggest that recall instructions, regardless of type or length, interfered with identification accuracy performance.

CHAPTER IV: CONCLUSION

General Discussion

The present study was designed to examine the effect of recall instructions on the VOE. By manipulating the type and length of recall instructions that participants were given after viewing a brief video of a burglary, I tested several predictions made by existing VOE theories and discussed avenues for future theory development. The results from Experiments 1 and 2 are discussed below with an emphasis on how they might contribute to a better understanding of the mechanisms driving the VOE.

Verbal Overshadowing Effect Size

The first aim of this study was to replicate the VOE and observe the size of the effect. The nature and reliability of the VOE has been discussed by researchers in the past (e.g., Alogna et al., 2014; Brown & Lloyd-Jones, 2003; Clifford, 2003; Dodson et al., 1997; Smith & Flowe, 2014), *and this paper attempted to first and foremost observe verbal overshadowing, and to do so using new stimuli (crime video and photo lineups) that have not been used in previous verbal overshadowing research.* Overall, VOEs were observed in both Experiments 1 and 2. Additionally, the verbal overshadowing effect sizes associated with recall instructions were large (ORs associated with observed verbal overshadowing ranged between 2.61 and 9.86); suggesting that the VOE is a reliable phenomenon and can be replicated across studies and using new test materials and stimuli.

Identification Accuracy

A second aim of the current study was to determine whether VOEs result from a reduction in discriminability, response criteria, or both. Results suggest that the VOE was a result of both a decrease in discriminability and more conservative response criteria. The results suggest that being asked to engage in recall decreased participants' ability to discriminate between the face that they saw in the video and the other faces in the photo lineup. The results also suggest that being asked to recall the suspect encouraged participants to adopt a more conservative response criteria. This latter finding (more conservative responses on the lineup identification after describing a face) supports the response CS account. However, this finding does not explain *why* providing a description of a person would make participants more conservative responders. This was a point Mickes and Wixted (2015) discussed in their paper regarding underlying mechanisms driving the VOE. Nonetheless, the observation that the VOE seems to be a result of a decrease in both discriminability and response criteria is a potentially important observation regarding future theory development.

Recall Instructions

The third aim of this study was to examine the influence of whether different recall instructions influence the verbal overshadowing effect in different ways. The addition, in the present study, of the 'general recall' instruction condition offered an opportunity to test implications of both the TIPS and the RBI theories. For instance, if participants who provided a general description of the suspect had better identification accuracy compared to participants who provide a detailed description of the face, results would have supported the TIPS account (e.g., Meissner & Brigham, 2001; Schooler, 2002). Additionally, if participants who described the facial features of the person had poorer identification accuracy compared to participants who provided a general description of the burglar, and if participants who described the facial features of the person had more incorrect details in their description compared to participants who provide a general description, then results would have supported the RBI account (e.g., Finger & Pezdek, 1999; Meissner at al., 2001).

In Experiment 1, identification accuracy was compared between participants who were asked to provide a general description of a suspect and participants who were asked to provide a detailed description of a suspect’s facial features. Previous verbal overshadowing studies have only asked participants to describe the face of an encoded target (e.g., Meissner et al., 2001; Schooler & Engstler-Schooler, 1990; Smith & Flowe, 2015). There are no published studies that have asked participants to engage in a free recall of an encoded target. Results revealed verbal overshadowing in the ‘face recall’ instruction condition compared to the ‘no recall’ instruction condition; verbal overshadowing was not found in the ‘general recall’ instruction condition compared to the ‘no recall’ instruction condition.

In order to further examine why asking participants to provide a detailed description of a suspect’s face would result in verbal overshadowing, both recall instruction type (general versus face) and recall instruction length (short versus long) were factors in Experiment 2. Previous studies manipulating various recall instructions often employed instructions that were long and detailed (see Table 4)—as a result, recall instruction length could have been a potential confound in those studies. If the VOE was observed in the ‘long recall’ instruction conditions in the present study, then support for that hypothesis would have been found. However, the results of the current study showed no effect of recall instruction length and no interaction with recall instruction type. Based on these results, I can confidently conclude that recall instruction length does not influence observed VOEs. Regarding the hit rate, verbal overshadowing was observed in all recall conditions. Regarding the false alarm rate, verbal overshadowing was observed in the 'face recall' instruction conditions (see Table 5). The results of Experiment 2 suggest that asking participants to describe the target, regardless of length of those recall instructions, result in poorer identification accuracy.

Discrepant Findings between Experiments 1 and 2

Hits. The results of Experiment 2 did not replicate the hit rate results of Experiment 1, despite the fact that the same 'general recall' instruction condition was used in both studies (the 'general short recall' instruction condition in Experiment 2 was identical to the 'general recall' instruction condition in Experiment 1). In Experiment 1, verbal overshadowing was not found when the 'general recall' instruction condition hit rate was compared to the 'no recall' instruction condition hit rate, but in Experiment 2, verbal overshadowing was found when the hit rates in those conditions were compared. One possible reason is that the VOE associated with providing a description of a suspect's general appearance may not be as reliable as the VOE associated with providing a description of a suspect's facial features. As a result, the discrepancy might simply reflect the lower power associated with the statistical test used in Experiment 1 compared to the power associated with the test in Experiment 2: it is possible that there was not enough statistical power to detect the effect of the 'general recall' instruction condition in Experiment 1 increasing the probability of a Type II error.

In order to address the discrepant findings in Experiments 1 and 2, I collapsed the data across the two studies and performed a *post-hoc* analysis comparing the 'general recall' and 'face recall' instruction conditions' hit rates to the 'no recall' instruction condition hit rate. I chose to include these three conditions in the analysis because these conditions represent direct replications across Experiments 1 and 2. That is, the same laboratory space, survey software, equipment, and college student population were used in both experiments, and there were no differences in the methodologies across experiments. A logistic regression consisting of only the ‘general short recall,’ ‘face long recall,’ and ‘no recall’ instruction conditions across both Experiments 1 and 2 were performed on hit rates and revealed a large VOE for both recall conditions, Wald $\chi^2(2) = 24.23$, $p > .001$. The odds of making a hit were significantly greater for participants in the ‘no recall’ condition than for participants in the ‘general short recall’ instruction condition, Wald $\chi^2(1) = 6.82$, $p = .009$, OR = 2.76 and the ‘face long recall’ instruction condition, Wald $\chi^2(1) = 23.99$, $p > .001$, OR = 6.98. These results strengthen the idea that engaging in recall of any kind might produce verbal overshadowing. The failure to find a VOE associated with the ‘general recall’ instruction condition in Experiment 1 might have been related to a relatively weak VOE associated with the ‘general recall’ instruction condition and the relatively low power of the statistical test in Experiment 1 compared to the relatively high power of the test in Experiment 2.

The previously reported discriminability (*d’*) analyses comparing the ‘no recall’ instruction condition with the ‘general recall’ instruction condition in Experiment 1 and in Experiment 2 further support the idea that there is a VOE effect associated with ‘general recall’ instruction condition. In both experiments, their respective *d’* values suggest that ‘general recall’ instructions interfered with discriminability. Further, the *d’* for the ‘no recall’ and ‘general recall’ comparison was actually larger in Experiment 1 (*d’* = 2.19) than in Experiment 2, (*d’* = 1.50). These analyses along with the previous analysis of hit rates provide support for a VOE associated with the ‘general recall’ instruction condition that is directly related to discriminability.

False alarms. While the hit rate results were inconsistent in the experiments, the results regarding false alarm rates were replicated. Across both Experiments 1 and 2, results show verbal overshadowing when participants are asked to provide descriptions of a target's facial features and not when asked to provide descriptions of a target's general appearance. In order to address this difference in hit and false alarm rate results, I turn to the results of the measures of discriminability (*d*') and response criteria (*C*) results (see Tables 1 and 5). These results combine the hit rate and false alarm rate data and demonstrate verbal overshadowing in all recall conditions compared to the 'no recall' instruction condition.

d' and C. Across Experiments 1 and 2, the joint hit and false alarm rate results reflected in the measures of *d'* and *C* show that participants who engaged in recall of the suspect's facial features had lowest discriminability and more conservative response criteria, followed by participants who engaged in recall of the suspect's general appearance. More importantly, however, both recall of the suspect's general appearance and recall of the suspect's facial features resulted in lower discriminability and more conservative response criteria than participants who did not provide a description. The discriminability and response criteria results reveal the influence of recall instructions on hit and false alarm rates jointly, and according to these results, verbal overshadowing was observed when participants provided a description of a suspect's general appearance and when participants provided a description of a suspect's facial features.

Description Accuracy

A fourth aim of the present study was to examine the relationship between description accuracy and identification accuracy. It may seem intuitive that accurate descriptions would be associated with accurate identifications. In fact, some studies have observed correlations between description accuracy and identification accuracy (Goldstein et al., 1979; Howells, 1938; Wolfskiel & Brigham, 1985). However, the majority of research regarding the description accuracy and identification accuracy relationship suggests that there is no association between description accuracy and identification accuracy (Finger & Pezdek, 1999; Meissner, 2002; Meissner et al., 2001; Smith & Flowe, 2014). Regardless of the research suggesting little association between descriptions and identifications, the present study investigated the relationship between description accuracy and identification accuracy as a means to better understand mechanisms driving the VOE.

While the current study revealed relationships between both recall instructions and description accuracy and recall instructions and identification accuracy, a meaningful relationship between description accuracy and identification accuracy was not found. For instance, in Experiment 1 in the 'face recall' instruction condition, a higher number of descriptors was associated with more false alarms. But, in Experiment 2 these same results were not replicated. This is only one of many inconsistencies in the description accuracy and identification accuracy across Experiments 1 and 2. These variations in the results make it difficult to interpret any relationship that might exist between description accuracy and identification accuracy in the present study. While I was unable to find a connection between description accuracy and identification accuracy in the present study, I believe that assessing the relationship between description accuracy and identification accuracy could be important for understanding why the VOE occurs. As such, I would like to address an important issue with regard to assessing such relationship.

Researchers largely attempt to find support for the various theories by simply correlating continuous measures of description accuracy (e.g., correct descriptors, incorrect descriptors) with dichotomous measures of identification accuracy (e.g., hits, false alarms). This is how the relationship was assessed in the present study. While correlating description accuracy with identification accuracy is typical in previous research on the VOE, both the RBI and TIPS accounts of verbal overshadowing are basically mediation theories, thus, correlation analysis is not an adequate way to test those theories. According to the RBI and TIPS accounts, various measures of description accuracy, might be mediating the relationship between recall instructions and identification accuracy. In the case of the RBI account, researchers believe that the accuracy of the description is responsible for the relationship between various recall instructions and identification accuracy. Particularly, the relationship between recall instructions and hits might be mediated by correct descriptors in a positive direction, and likewise, the relationship between recall instructions and false alarms might be mediated by incorrect descriptors in a positive direction. Regarding the TIPS account, some researchers (e.g., Schooler, 2002; Schooler & Engstler-Schooler, 1990) believe that the relationship between recall instructions and identification accuracy might be mediated by a type of processing which might be represented by number of descriptors: specifically, that number of descriptors would mediate recall instructions on hits in a negative direction, and that the number of descriptors would mediate recall instruction on false alarms in a positive direction. Because the existing VOE theories make these specific hypotheses, I believe that using mediation analysis is the most appropriate method for testing the RBI and TIPS accounts and that correlation analysis may not adequately test these theories.

While I think that mediation analysis might be the ideal way to examine the mediating effects of description accuracy, another concern remains: in order to test whether measures of description accuracy mediate the relationship between recall instructions and identification accuracy, the measures of description accuracy must contain a relatively large range of values. The reason I chose not to report the results regarding the relationship between description accuracy and identification accuracy using mediation analysis in the present study is because the description accuracy frequencies were small: *M*s for correct descriptors and incorrect descriptors ranged from 0.25 to 5.98 and *SD*s ranging from 0.43 to 2.34. Given the low frequency of occurrence of these responses, an analysis testing for mediated effects of description accuracy would, simply, not be interpretable. For instance, collapsed across recall instruction condition, the mean for incorrect descriptors in Experiment 1 was $M = 0.39$ with a $SD = 0.58$. The frequencies of correct and incorrect descriptors reported in the present study are similar to those reported in previous studies (e.g., Meissner, 2002; Meissner et al., 2001; Smith & Flowe, 2015). If studies' measures of description accuracy were larger and possessed more variability, then conducting mediation analyses would be the most informative way to examine the extent to which description accuracy mediates the relationship between recall instructions and identification accuracy.

Lineup Identification Response Time

The fifth and final aim of this study was to examine the relationship between lineup identification response time and identification accuracy. Previous research on lineup identification response time and identification accuracy suggests that faster lineup identification response times are associated with an increase in correct identifications from lineups, while slower response times are associated with an increase in incorrect identifications (Brewer, et al., 2006; Bruer & Pozzulo, 2014; Dunning & Perretta, 2002; Sauerland & Sporer, 2009; Sporer, 1993; Weber et al., 2004). Additionally, research conducted to specifically examine lineup identification response time and identification accuracy as it pertains to the VOE found that forcing participants to respond quickly during the lineup identification task resulted in improved identification accuracy as opposed to not being forced to respond quickly during a lineup identification (Schooler & Englster-Schooler, 1990, Exp 6). While participants in the present study were not forced to respond on the lineup identification task within any time frame, participants' lineup identification responses were timed and recorded in the present study. Lineup identification response times were examined as a function of recall instructions, and the relationship between lineup identification response time and identification accuracy was assessed.

In Experiment 1, participants in the 'no recall' instruction condition made faster lineup identifications compared to the 'general recall' and 'face recall' instruction conditions, and the 'general recall' instruction condition made faster lineup identifications compared to the 'face recall' instruction condition; see Table1. When the relationship between lineup identification response time and identification accuracy was assessed, various relationships were found. In the 'no recall' instruction condition, faster response times were associated with more hits and slower response times were associated with more false alarms. In the 'general recall' instruction condition, faster response times were also found to be associated with more hits. In contrast, in the 'face recall' instruction condition, faster response times were associated with more false alarms. The results observed in the 'no recall' and 'general recall' instruction conditions are consistent with the previous research suggesting that faster response times are associated with better accuracy and slower response times are associated with poorer accuracy (Brewer, et al., 2006; Bruer & Pozzulo, 2014; Dunning & Perretta, 2002; Sauerland & Sporer, 2009; Schooler & Englster-Schooler, 1990; Sporer, 1993; Weber et al., 2004). However, the finding in the 'face recall' instruction condition is not consistent with the literature: in fact, the finding that faster response times were associated with more false alarms was surprising. It is not entirely clear why asking participants to describe the face of a target would influence performance on the target absent lineup in this fashion, however, it is possible that participants in the 'face recall' instruction condition experienced greater difficulty in their ability to distinguish between the suspect's face from the other faces in the lineup, and decided to make an immediate lineup identification rather than take a longer amount of time to make an identification. While this is one possibility, previous research suggests that difficulty in making discrimination judgments are typically associated with longer response time measures.

In Experiment 2, participants in the 'no recall' instruction condition made faster lineup identification response times compared to participants in the 'general short,' 'general long,' 'face short,' and 'face long' recall instruction conditions, see Table 5. This finding was expected and consistent with what was found in Experiment 1. However, when assessing the relationship between identification response time and identification accuracy, not all results were replicated. For instance, overall, across all recall instruction conditions, faster response times were associated with more hits. In the 'no recall' instruction condition, faster response times were also associated with more hits. Unlike Experiment 1, there was no relationship between response times and hits or false alarms in the 'general short' recall instruction condition. Additionally, there was not a relationship between identification response times and hits and false alarms in the 'general long' recall instruction condition. In the 'face short' recall instruction condition, faster response times were associated with more hits. In the 'face long' recall instruction condition, faster response times were associated with more hits. Unlike Experiment 1, there was no correlation between response times and false alarms, and the finding that faster response times were associated with more false alarms was not replicated in Experiment 2.

While the current study does not provide a clear story of how lineup identification response time can be beneficial in understanding more about the VOE phenomenon, the relationship between identification response time and identification accuracy should not be discounted as a diagnostic tool regarding verbal overshadowing. Differences found in the relationship between lineup identification response time and identification accuracy for each recall instruction condition might be informative for future theory development regarding the VOE.

Theoretical Implications

The present study's results provide support for various aspects of the current VOE theories. The CS account proposes that verbalization primarily affects the decision strategy, or criterion, that participants adopt during a recognition test. This account suggests that when participants find it difficult to describe a suspect, they feel uncertain about the accuracy and/or completeness of their description. As a result, participants may doubt the strength of their memory and adopt a more conservative decision standard during a recognition test. Because target-absent lineups were used in the present study and false alarms rates were obtained, I was able to test the CS account. The present study revealed partial support for the CS account. The response criteria (*C*) results showed that participants in the present study who were either asked to provide a description of the suspect's face or suspect's general appearance, had a more conservative response criteria compared to participants who did not describe the suspect. The results are consistent with previous studies that have also found support for the CS account (Clare & Lewandowsky, 2004; Smith & Flowe, 2014; Sauerland, Holub, & Sporer, 2004). Consistent with the CS account, results suggest that these participants were uncertain of their descriptions and, as a result, influenced their lineup identification strategy.

The RBI account focuses on how recall instructions impact memory. The theory suggests that when participants engage in a detailed recall task (e.g., describe the facial features of a suspect), they are more likely to have errors in their descriptions (Experiment 1) and provide more subjective descriptors (Experiment 2). These descriptions, in turn, might affect the accuracy of participant's subsequent identification. Partial support for the RBI account was found. More specifically, across two experiments, participants in the 'face recall' instruction conditions provided more incorrect descriptors and subjective descriptors than participants in the 'general recall' instruction condition. However, participants in the 'face recall' conditions did not perform worse on the identification task compared to participants in the 'general recall' conditions. Overall, these results are only partially consistent with previous work that has found support for the RBI account (Finger & Pezdek, 1999; Meissner, 2002; Meissner et al., 2001; Smith & Flowe, 2014).

While the present study's results speak to the CS and RBI accounts, I think the current study's results mainly support the TIPS account. At the moment, the TIPS account is the only processing account put forth to explain verbal overshadowing. The TIPS account proposes that accurate facial identifications are facilitated by configural processing, whereas providing verbal descriptions is primarily a featural processing task: verbal overshadowing is produced when the configural process that is used during the encoding and recognition of a face is interrupted by a featural process that is used during recall. I believe the present study's results support the part of the TIPS account that suggests that participants experience a shift from configural to featural processing when asked to engage in a verbalization task. Additionally, I think the study's results extend this finding and appear to indicate that both types of recall instructions—recall instructions regarding the general appearance of a person and recall instructions regarding the facial features of a person— initiate a switch from configural to featural processing.

I think that the results of the present study contribute to a better understanding of why verbal overshadowing occurs in two important ways. First, the findings that verbal overshadowing was observed in both the 'general recall' and 'face recall' instruction conditions suggest that the type of information that participants are asked to engage in during the recall task may not influence verbal overshadowing differently. Results appear to show that verbal overshadowing might occur no matter what type of recall people are asked to engage in. Results are consistent with previous research that has also found impaired recognition performance after describing a face (e.g., Finger & Pezdek, 1991; Meissner, 2002, Meissner et al., 2001; Schooler & Englster-Schooler, 1990), or engaging in other proposed featural processing-type tasks (e.g., Brown & Lloyd-Jones, 2003; Dodson et al. 1997). Second, results indicating that recall instruction length did not seem to influence the VOE suggest that the impaired identification performance was not a result of participants trying to remember the recall instructions while simultaneously trying to engage in recall. In Experiment 2, the present study examined recall instruction length as a potential confound across previous verbal overshadowing studies: that is, verbal overshadowing researchers employ long instructions in their studies, and the influence of the length of the recall instructions on verbal overshadowing had never been examined as a potential confound. The results of Experiment 2 appear to eliminate that as a possible explanation for the VOE.

Policy Implications

The current research has potentially important applied implications for the legal system regarding how police administer lineups. The current study addressed the guidelines written by the National Institute of Justice which state that law enforcement should "…encourage the witness to report *all* [my italics] details, even if they seem trivial" might hinder identification accuracy (*Eyewitness Evidence: A Guide for Law Enforcement*, 1999, p. 22). That is, the present study questioned whether, asking an "all" question might reduce identification accuracy. However, the results of the present study did not find an effect of this type of instruction on identification accuracy. In fact, the findings suggest that asking eyewitnesses to engage in a description of any type or length regarding a suspect might impair later identification accuracy of that suspect.

Future Research

I would emphasize four issues as a guide for future research on the VOE. First, I think future research should be done to understand why describing the face of a target as well as describing the general appearance of a target interferes with identification accuracy performance. That is, if the difference in types of processes induced when describing a face than when describing the general appearance of a person does not result in differences in identification accuracy, then what types of processes are underlying the verbal overshadowing phenomenon? I think the current study's results demonstrate that recall of a target's general appearance and recall of target's facial features share a common processing mechanism that results in verbal overshadowing. Future research might use this information to identify other types of tasks that are associated with processes that also interfere with identification performance. Such research will be useful in determining more definitively the types of cognitive mechanism that are responsible for reduced identification accuracy.

Second, future research needs to address the issue regarding the relationship between description accuracy frequency and identification accuracy. Assessing the relationship between description accuracy and identification accuracy seems necessary to establish whether recall and recognition memory use common memory mechanisms. But, until a better way to assess the relationship between the two variables is established, such as mediation, interpretation of any description accuracy and identification accuracy relationship remains uncertain.

Third, I believe that research on the relationship between lineup identification response time and identification accuracy might provide insight regarding the VOE and future theory development. Specifically, response times have long been associated with decision-making accuracy (e.g., Brewer, et al., 2006; Dunning & Perretta, 2002; Weber et al., 2004). Thus, verbal overshadowing researchers might use identification response time measures to help explain their findings. Previous verbal overshadowing researchers seem to ignore and neglect reporting any effects of response time on identification accuracy in the VOE literature. I recommend that verbal overshadowing researchers both record and report response times in their future studies.

The last recommendation for future research that I emphasize in the present paper, is for researchers to investigate the VOE via a cognitive psychological approach. The original studies that demonstrated the VOE were forensic psychology studies. Schooler and Engstler-Schooler's use of forensic-type materials (e.g., encoding of a criminal, recognition of the criminal from a photo lineup) allowed forensic and other applied researchers to study the VOE in an applied area (e.g. Finger & Pezdek, 1999; Smith & Flowe, 2014; Meissner, 2002; Meissner & Brigham, 2001). As a result, those studies were successful in establishing that VOE occurred in the context of eyewitness memory but failed to explain why the VOE occurred. In order to understand *why* the VOE occurs, and what cognitive mechanisms are responsible for verbal overshadowing, I recommend that future research approach the VOE as an issue in cognitive psychology. Specifically, research conducted on the VOE without the use of crime videos, lineups, and other applied materials might be done to establish the VOE as an issue addressable outside of those applied forensic psychology parameters.

Conclusion

In conclusion, the present study was designed to better understand why verbal overshadowing occurs and to assess the magnitude of the effect. In order to contribute to a better understanding of the VOE phenomena, the present study examined the influence of different recall instructions on the VOE. Results reveal that asking participants to provide a description of a suspect resulted in verbal overshadowing. Regarding VOE theories, the finding that providing a description of a person, regardless of recall instruction length and recall instruction type, interferes with subsequent identification accuracy contributes to future theory development for the VOE. Specifically, results seem to support a processing account for verbal overshadowing. In the present paper, I offer several future avenues of research regarding the VOE including more research on recall instructions, the relationship between description accuracy and identification accuracy, and identification response time. The need for research regarding the VOE to be conducted in cognitive psychology specifically is also addressed. Lastly, results also have applied implications for how law enforcement administer lineups. According to the results, asking eyewitnesses to provide a description of any type could encourage the eyewitness to engender a more conservative lineup identification strategy and be less likely to identify the suspect correctly. On the other hand, not asking an eyewitness to provide a description could encourage the eyewitness to adopt a more liberal approach when making a lineup identification and be more likely to identify a face of an innocent suspect as well as the face of the suspect.

BIBLIOGRAPHY

BIBLIOGRAPHY

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahnik, S., Birch, S., Birt, A. R., . . . Zwaan, R. A. (2014). Registered replication report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*, 556–578.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30*, 31-50.

Brown, C., & Lloyd-Jones, T. J. (2003). Verbal overshadowing of multiple face and car recognition: Effects of within- versus across-category verbal descriptions. *Applied Cognitive Psychology*, *17*, 183-201.

Bruer, K., & Pozzulo, J. D. (2014). Does speed indicate lineup identification accuracy? Examining children's and adults' reaction time. *American Journal of Forensic Psychology*, *32*, 43-61.

Chandler, C. C. (1993). Accessing related events increases retroactive interference in a matching recognition rest. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 967-974.

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 739-755. DOI: 10.1037/0278-7393.30.4.739

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629-654.

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 395-424.

Clifford, B. R. (2003). The verbal overshadowing effect: In search of a chimera. In M Vanderhallen, G. Verwaeke, P. J. van Koppen, & J. Goethals (Eds.), Much ado about crime: *Chapters on psychology and law* (pp. 151-162). Brussels: Politieia.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87-114.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., and McGorty, E. K. A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28*, 687-706.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107–117. DOI: 10.1037/0096-3445.115.2.107

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

Dunning, D. & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12-s rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology*, *87*, 951-962.

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*, 818-835.

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

Francis, G. (2012). Too good to be true: Publication bias to two prominent studies from experimental psychology. *Psychonomic Bulletin & Review, 19*, 151-156.

Gobet, F., & Clarkson, G. (2004). Chunk in expert memory: evidence for the magical number four…or is it two? *Memory, 12*, 732-747.

Goldstein, A. G., Johnson, K. S., & Chance, J. (1979). Does fluency of face description imply superior face recognition? *Bulletin of the Psychnomic Society, 13,* 15-18.

Howells, T. H. (1938). A study of ability to recognize faces. *The Journal Of Abnormal And Social Psychology, 33*, 124-127. doi:10.1037/h0062942

Kaye, D. H. (1986). Quantifying probative value. *Boston University Law Review, 66*, 71-81.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

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

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

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29*, 176–186. DOI: 10.3758/BF03195751

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

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristics analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous vs. sequential lineups. *Journal of Experimental Psychology: Applied*, *18*, 361-376. DOI: 10.1037/a0030609

Mickes, L., & Wixted, J. T. (2015). On the applied implications of the “verbal overshadowing effect.” *Perspectives on Psychological Science, 10*, 400-403. DOI: 10.1177/1745691615576762

National Institute of Justice, Eyewitness Evidence: A Guideline for Law Enforcement (1999).

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

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior, 36*, 247-255.

Sauerland, M., Holub, F. E., & Sporer, S. L. (2008). Person descriptions and person identifications: Verbal overshadowing or recognition criterion shift? *European Journal of Cognitive Psychology*, *20*, 497–528. DOI: 10.1080/09541440701728417

Sauerland, M., & Sporer, S. L. (2009). Fast and confidence: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, *15*, 46-62. DOI: 10.1037/a0014560

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology*, *16*, 989–997. DOI: 10.1002/acp.930

Schooler, J. W. (2011) Unpublished results hide the decline effect. *Nature*, 470, 437.

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

Smith, H. J. M., & Flowe, H. D. (2015). ROC analysis of the verbal overshadowing effect: Testing effect of verbalization on memory sensitivity. *Applied Cognitive Psychology, 29*, 159-168. DOI: 10.1002/acp.3096

Smith, S. M., Lindsay, R. C. L., Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology, 85*, 542-550.

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22*, 157-180.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 87*, 22-33.

Sporer, S. L. (1996). Describing others: Psychological issues. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), Psychological issues in eyewitness identification (pp. 53-86). Hillsdale, NJ: Lawrence Erlbaum

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413-424.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, *88*, 776-784.

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification and response latency. The unruly 10-12 s rule. *Journal of Experimental Psychology: Applied*, *10*, 139-147.

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon "probative value" and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science, 7*, 275-278.

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review, 121*, 262-276.

Wolfskiel, M. P., & Brigham, J. C. (1985). The relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology, 70*, 547-555. DOI:10.1037/0021-9010.70.3.547

LIST OF APPENDICES

APPENDIX A

Crime Video

Photo Lineups

Target Present Lineup

Target Absent Lineup

Demographic Information Questionnaire

1. What is your age?

2. What is your sex?

   Male
   Female

3. How do you describe yourself? (Please check the one option that best describes you.)

   Black or African American
   Hispanic or Latina
   Non-Hispanic White
   Other [fill in]

4. What is your class standing?

   Freshman
   Sophomore
   Junior
   Senior
   Graduate student [fill in what year]

APPENDIX B

Table 1. Experiment 1 hit rates, false alarm (FA) rates, discriminability (*d'*), response criteria (*C*), and lineup identification response time (RT) for each recall instruction condition

Recall instruction condition	Hit rate	FA rate	*d'*	*C*	RT
No recall	0.84	0.23	2.75	0.38	10.07
General recall	0.67	0.24	2.19	0.66	15.35
Face recall	0.34	0.59	0.87	0.85	18.81

Table 2. Experiment 1 descriptive results for description accuracy measures across each recall instruction condition

Description accuracy measure	Recall instruction condition	*M*	*SD*	Min	Max
Correct descriptors	General recall	*2.94	1.10	1	6
	Face recall	2.48	1.45	0	6
Incorrect descriptors	General recall	**0.25	0.43	0	1
	Face recall	0.52	0.73	0	3
Subjective descriptors	General recall	**5.61	2.47	1	13
	Face recall	4.35	2.51	0	11
Number of descriptors	General recall	**8.80	2.63	3	15
	Face recall	7.35	2.91	2	14

Note. *$p < .05$, **$p < .01$

Table 3. Experiment 1 lineup identification response times for hits and false alarms (FAs) across recall conditions (overall) and for each recall instruction condition

Recall instruction condition	Hits		FAs	
	Yes	No	Yes	No
No recall	***8.22	18.30	*8.95	14.77
General recall	14.31	18.38	*13.39	20.05
Face recall	21.28	15.33	***24.60	16.77
Overall	12.61	16.73	14.27	17.03

Note. *p < .05, ***p < .001

Table 4. A comparison of our study's 'face recall' instruction condition to other studies' similar recall instructions

Author(s) Study's recall instructions condition	Recall instructions	Number of words
Baker & Reysen 'Face recall' instructions	Please describe the face you saw in the video. Your task is to describe the person in such a way that your description would aid someone else in attempting to identify the person. Your description should focus on facial features. Write about the shape and size of the eyes, eyebrows, nose, ears, mouth, chin, etc. Try not to leave out any details about the face even if you think they are not important.	73
Finger & Pezdek (1991) Exp 1 'Standard interview' instructions	Please try to describe the person you were shown at the beginning of the experiment. Specifically, your task is to describe the person in such a way that your description would aid someone else in attempting to identify the person. Your description should therefore focus primarily on physical features. You might begin for example, by describing the person's sex, race, age, hair color and style, facial features, and so on. Try to be as complete as possible, but do not guess at things about which you are uncertain.	88
Meissner (2002) Exp 1 'Free recall' instructions	In the spaces below, please describe the face you saw in the slide. Use the lines below to provide details about what the face looked like. You should attempt to describe the person in sufficient detail such that someone else could identify him on the basis of the description. As describing a face is often a difficult task, it is important that you concentrate and stay focused for the next few minutes.	72
Smith & Flowe (2014) 'Standard recall' instructions	In the box below, please describe the face that you saw in the video. Your task is to describe the person in such a way that your description would aid someone else in attempting to identify the person. Your description should focus on facial features. Write about the shape and size of the eyes, eyebrows, nose, ears, mouth, chin, etc.	60

Note. The first column lists the authors of the studies that possessed recall instructions similar to the present study's 'face recall' instruction condition. The first column also provides the name of the recall instruction condition that those authors used to label that recall condition.

Table 5. Experiment 2 hit rates, false alarm (FA) rates, discriminability (*d'*), response criteria (*C*), and lineup identification response time (RT) for each recall instruction condition

Recall instruction condition	Hit rate	FA rate	*d'*	*C*	RT
No recall	0.80	0.32	2.49	0.41	8.82
General short recall	0.58	0.38	1.76	0.68	15.36
General long recall	0.49	0.43	1.45	0.75	16.86
Face short recall	0.48	0.65	1.15	0.65	16.74
Face long recall	0.44	0.67	1.08	0.69	15.77

Table 6. Experiment 2 descriptive results for description accuracy measures across each recall instruction condition

Description accuracy measure	Recall instruction condition	*M*	*SD*	Min	Max
Correct descriptors	Recall instruction type				
	General recall	***5.98	2.34	1	11
	Face recall	3.63	1.75	0	8
	Recall instruction length				
	Short recall	***4.05	2.25	0	10
	Long recall	5.56	2.26	1	11
Incorrect descriptors	Recall instruction type				
	General recall	**0.66	0.78	0	4
	Face recall	0.45	0.6	0	2
	Recall instruction length				
	Short recall	***0.37	0.6	0	3
	Long recall	0.74	0.75	0	4
Subjective descriptors	Recall instruction type				
	General recall	***1.27	1.37	0	7
	Face recall	2.57	1.91	0	9
	Recall instruction length				
	Short recall	***1.50	1.43	0	7
	Long recall	2.35	2	0	9
Number of descriptors	Recall instruction type				
	General recall	***7.90	2.78	1	14
	Face recall	6.65	2.95	1	16
	Recall instruction length				
	Short recall	***5.92	2.52	1	13
	Long recall	8.65	2.68	2	16

Note. **$p < .01$, ***$p < .001$

Table 7. Experiment 2 lineup identification response times for hits and false alarms (FAs) across recall conditions (overall) and for each recall instruction condition

Recall instruction condition	Hits		FAs	
	Yes	No	Yes	No
No recall	***7.03	12.05	9.28	10.22
General short recall	12.76	18.08	15.54	16.13
General long recall	14.48	17.70	17.01	18.37
Face short recall	*14.03	18.20	17.81	16.93
Face long recall	**12.95	17.57	15.66	16.20
Overall	***11.70	17.37	14.49	15.94

Note. *$p < .05$, **$p < .01$, ***$p < .001$

VITA

Melissa A. Baker

Peabody Hall, Room 310
Department of Psychology
University of Mississippi
Oxford, MS 38677

email: mbaker5@go.olemiss.edu

PREVIOUS EMPLOYMENT

2015-2016 Adjunct Instructor, General Experimental Psychology
Appalachian State University, Boone, North Carolina

EDUCATION

2016-present Ph.D. in Experimental Psychology
University of Mississippi, Oxford, Mississippi

2013-2015 M.A. in General Experimental Psychology
Appalachian State University, Boone, North Carolina

2012-2013 Certification Abroad in Research Psychology
University of New South Wales, Sydney, Australia

2009-2012 B.S. in Psychology
Minor in Criminal Justice
Appalachian State University, Boone, North Carolina

RESEARCH INTERESTS

My research focuses on applied issues in cognitive psychology. My research interests include:

- Eyewitness memory (i.e., eyewitness identification accuracy of suspects from lineups)
- The effect of recall memory on recognition memory (i.e., verbal overshadowing/facilitation effects)
- Contextual-framing and -priming on individual's judgments of ambiguous situations (i.e., the public's perceptions of police officer-civilian confrontation videos)
- Research methodologies (i.e., crowdsourcing) and quantitative methods (i.e., Bayesian approaches) used in forensic contexts

PEER REVIEWED PUBLICATIONS

Baker, M. A., & Reysen, M. B. (2019). The influence of pre-video information and appraisal time on judgments of police officer use of excessive force. *Journal of Police and Criminal Psychology*. DOI: 10.1007/s11896-018-9295-8

Reysen, M. B., Bliss, H., & Baker, M. A. (2018). Survival processing eliminates collaborative inhibition. *Quarterly Journal of Experimental Psychology, 71(6),* 1-25. DOI: 10.1080/17470218.1318408

Baker, M. A., & Bacharach, V. R. (2017). Police officer-civilian confrontations caught on camera: The influence of contextual frames on judgments of excessive force. *American Journal of Criminal Justice, 42(4),* 683-697. DOI: 10.1007/s12103-017-9387-5

Baker, M. A., Fox, P., & Wingrove, T. (2016). Crowdsourcing as a forensic psychology research tool. *The American Journal of Forensic Psychology, 34(1),* 37-50. DOI: 10.13140/RG.2.1.11308.3129

PAPER PRESENTATIONS (*ASU Undergraduate Student)

Baker, M. A., & Bacharach, V. R. (2016, June). Better late than never: The effect of time of recall on eyewitness identification accuracy of criminals. Paper presented at Law and Society, New Orleans, LA.

Baker, M. A., & Bacharach, V. R. (2016, March). The effects of warnings on perceptions of police officer use of excessive force. Paper presented at the American Psychology-Law Society, Atlanta, GA.

Baker, M. A., & Bacharach, V. R. (2016, March). Factors that influence perceptions of police officer use of excessive force. Paper presented at the American Psychology-Law Society, Atlanta, GA.

Baker, M. A., Fox, P. A., & Wingrove, T. (2015, June). Own-gender bias in lineup identifications. Paper presented at the Society for the Psychological Study of Social Issues, Washington, D.C.

Anderson, A., Baker, M. A., Mackinnon, A., Wingrove, T., Fox, P. A., & Ferrone, C*. (2015, March). Gendered perceptions of teacher-student sexual relationships. Paper presented at the American Psychology-Law Society, San Diego, CA.

POSTER PRESENTATIONS (ǂUM Undergraduate Student, *ASU Undergraduate Student)

Dean, M.ǂ, Baker, M. A., & Reysen, M. B. (2019, March). Testing the response criteria shift account for the verbal overshadowing effect. Poster presented at the American Psychology-Law Society, Portland, OR.

Baker, M. A., & Reysen, M. B. (2018, March). The influence of recall instructions of the verbal overshadowing effect. Poster presented at the American Psychology-Law Society, Memphis, TN.

Baker, M. A., & Johnson, M. (2018, March). Urban, suburban, and rural attitudes toward police brutality. Poster presented at the 64th Meeting of the Southern Psychological Association, Charleston, SC.

Baker, M. A. (2017, September). Community attitudes towards police misuse of authority. Poster presented at the Southern Criminal Justice Association, New Orleans, LA.

Baker, M. A., & Reysen, M. B. (2017, August). The VOE: A problem with discriminability, response criteria, or both? Poster presented at the American Psychological Association, Washington, D.C.

Baker, M. A., Johnson, M., & Castiglia-Marelli, C.* (2017, March). The influence of the media on public attitudes towards police. Poster presented at the 63rd Meeting of the Southeastern Psychological Association, Atlanta, GA.

Baker, M. A. (2016, September). Relationship between residential living area and perceptions of police prejudice towards minorities. Poster presented at the Southern Criminal Justice Association, Savannah, GA.

Baker, M. A., Fox, P. A., & Wingrove, T. (2016, September). The effect of eyewitness gender on free recall of criminals' physical descriptions. Poster presented at the Southern Criminal Justice Association, Savannah, GA.

Baker, M. A., Fox, P. A., & Wingrove, T. (2016, August). Using crowdsourced samples in forensic psychological research. Poster presented at the American Psychology Association, Denver, CO.

Johnson, M., Castiglia-Marelli, C*., Baker, M. A., & Bacharach, V. R. (2015, November). Prior arrest history by age interaction on perceptions of police officer use of excessive force. Poster presented at the North Carolina Psychological Foundation, Durham, NC.

Baker, M. A., Fox, P. A., & Wingrove, T. (2015, September). Gender differences in recall of criminals. Poster presented at the Southern Criminal Justice Association, Charleston, SC.

Baker, M. A., Fox, P. A., & Wingrove, T. (2015, May). Effect of recall on recognition of criminals. Poster presented at the Association for Psychological Science, New York, NY.

Krauza, R. M.*, Baker, M. A., Fox, P. A., & Wingrove, T. (2015, April). Own-gender bias in descriptions of criminals. Poster presented at the North Carolina Psychology Association Convention, Chapel Hill, NC.

Poff, B.*, Baker, M. A., Fox, P. A., & Wingrove, T. (2015, March). Gender bias in criminal descriptions. Poster presented at the Academy of Criminal Justice Sciences 52nd Annual Meeting, Orlando, FL.

Poff, B.*, Baker, M. A., Fox, P. A., & Wingrove, T. (2014, May). Effect of gender bias and verbal ability on accuracy of describing a criminal offender. Poster presented at the 17th Annual Celebration of Student Research and Creative Endeavors, Boone, NC.

Aguliar, D.*, Baker, M. A., Poff, B.*, Fox, P. A., Wingrove, T., & Anderson, A. (2013, November). Immigrant bias of jurors. Poster presented at the State of North Carolina Undergraduate Research and Creativity Symposium, Charlotte, NC.

Steele, K. M., Baker, M. A., Kinura, N., Gray, J., Strickland, E., & Barrett, P. (2012, November). Can a Stroop procedure induce the Mehta and Zhu (2009) color effect? Poster presented at the 53rd meeting of the Psychonomic Society, Minneapolis, MN.

INVITED TALKS

Baker, M. A., & Mackinnon, A. (2015, October). The psychology of individual biases in the legal system. Appalachian State University, Criminal Justice, meeting of the American Corrections Association, Boone, NC.

RESEARCH AWARDS

UM Graduate School Competitive Summer Research Assistantship Award, $3,000. Graduate School, University of Mississippi, Oxford, MS, April 2018.

Office of Student Research Grant, $200.00. Office of Student Research, Appalachian State University, Boone, NC, January 2014.

Cratis D. Williams Graduate Research Grant, $200.00. Office of Student Research, Appalachian State University, Boone, NC, January 2014.

WISE Research Award, $300.00. Psychology Department, Appalachian State University, Boone, NC, December 2013.

RESEARCH EXPERIENCE

2016-present Graduate Research Assistant, Psychology Department, University of Mississippi, Oxford, Mississippi.

2015-2016 Adjunct Instructor/Researcher, Psychology Department, Appalachian State University, Boone, North Carolina.

2013-2015 Graduate Research Assistant, Psychology Department, Appalachian State University, Boone, North Carolina.

2012-2013 Intern/Research Assistant, School of Psychology, University of New South Wales, Sydney, Australia.

2011-2012 Undergraduate Research Assistant, Psychology Department, Appalachian State University, Boone, North Carolina.

2011-2012 Undergraduate Research Assistant, Grandfather Home for Children, Banner Elk, North Carolina.

TEACHING EXPERIENCE

Appalachian State University, Boone, North Carolina
- PSY 1200: Introduction to Psychology, 2015-2016
- PSY 4217: Cognitive Psychology Lab, 2015-2016
- PSY 3100: Research Methods in Psychology Lab, 2014-2015

GRADUATE TEACHING ASSISTANT EXPERIENCE

University of Mississippi, Oxford, Mississippi
- PSY 704: Quantitative Methods in Psychology II (graduate course), 2019
- PSY 703: Quantitative Methods in Psychology I (graduate course), 2018
- PSY 392: Experimental Social Psychology, 2018
- PSY 319: Brain and Behavior, 2017-2018
- PSY 301: Developmental Psychology, 2019
- PSY 202: Statistics for Behavioral Sciences (tutor), 2017
- PSY 201: General Psychology, 2016

Appalachian State University, Boone, North Carolina
- PSY 3100: Research Methods in Psychology, 2014-2015
- PSY 5500: Forensic Psychology, 2013-2014

UNDERGRADUATE TEACHING ASSISTANT EXPERIENCE

Appalachian State University, Boone, North Carolina
- PSY 3100: Research Methods in Psychology, 2012
- BIO 1102: Biology in Society II, 2011
- BIO 1101: Biology in Society I, 2011