

2005

An assessment of eyewitness accuracy: The integration of suggestibility and misidentification

Karri Bonner
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Bonner, Karri, "An assessment of eyewitness accuracy: The integration of suggestibility and misidentification" (2005). *Graduate Theses, Dissertations, and Problem Reports*. 4135.
<https://researchrepository.wvu.edu/etd/4135>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

AN ASSESSMENT OF EYEWITNESS ACCURACY: THE INTEGRATION OF
SUGGESTIBILITY AND MISIDENTIFICATION

Karri Bonner, M.A.

Dissertation submitted to the Eberly College of Arts and Sciences at
West Virginia University in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
in
Life-Span Developmental Psychology

Matthew H. Scullin, Ph.D., Chair
Julie H. Patrick, Ph.D.
JoNell Strough, Ph.D.
William Fremouw, Ph.D.
Ryan Finkenbine, Ph.D.

Department of Psychology

Morgantown, West Virginia
2005

Keywords: Suggestibility, Assessment, Lineup Performance

Abstract

An Assessment of Eyewitness Accuracy: The Integration Of Suggestibility And Misidentification

Karri S. Bonner

The current study was an examination of eyewitness accuracy and how it relates to misidentification. A new scale to measure suggestibility in adults (The Video Suggestibility Scale for Adults, VSSA) was developed. This scale is modeled after the Gudjonsson Suggestibility Scales, which define suggestibility as a combination of Yield and Shift. Participants (N = 160) were administered the VSSA. Participants watched a video of two crimes and were subsequently interviewed either immediately or after a 5-8 day delay. The interview consisted of free recall, open-ended questions, and closed-ended questions suggestive questions. Participants also viewed several lineups to assess accuracy of identification of perpetrators. Results indicated that participants yielded to the suggestions made by the interviewer and shifted their answers in response to negative feedback. Timing of interview had an effect on Yield 1, Shift, and Total Suggestibility, with individuals who were interviewed immediately having lower Yield 1, Shift, and Total Suggestibility scores. Individuals who were interviewed after a delay were more likely to decline in the accuracy of their responses to probe questions than those who were interviewed immediately. Participants who were more suggestible were more likely to incorrectly select a perpetrator from a target-absent lineup, as were those who scored high on the Yield 2 subscale. These findings have important implications for how witnesses should be interviewed.

Acknowledgements

I would like to thank my committee members for all of the guidance throughout this project. These members include the chair, Dr. Matthew Scullin, as well as Drs. Julie Patrick, JoNell Strough, William Fremouw, and Ryan Finkenbine. I especially would like to acknowledge and thank my mentor and friend Dr. Stanley Cohen. Without him, I certainly would not be where I am today. He helped me to find the confidence that I needed and he was an inspiration to me both personally and professionally. I am deeply appreciative of all the wisdom that he shared with me about life as a graduate student, life after graduate school, and about life in general.

TABLE OF CONTENTS

Abstract.....	ii
Acknowledgements	iii
Table of contents	iv
List of Figures.....	v
List of Tables.....	vi
Introduction	1
Method.....	22
Results	29
Discussion.....	37
References	58
Appendices	64

LIST OF FIGURES

Figure 1.....	56
Figure 2.....	57

LIST OF TABLES

Table 1	49
Table 2	50
Table 3	51
Table 1	52
Table 2	53
Table 3	54

An Assessment of Eyewitness Accuracy: The Integration of Suggestibility and Identification

During criminal trials, the jury's decision to acquit or convict a defendant can often be the result of listening to eyewitness testimony. Unfortunately, eyewitness testimony is sometimes faulty and can place innocent people in prison. For example, Wells, Small, Penrod, Malpass, Fulero, and Brimacobe (1998) reported that of the 40 convictions overturned by DNA evidence, 36 (90%) of those innocent individuals were incarcerated based on faulty eyewitness testimony. Similarly, Scheck, Neufeld, and Dwyer (2000) found that out of another 60 convictions overturned, approximately 75% were overturned because of faulty eyewitness testimony.

There is a need to investigate this area and to find ways in which to evaluate eyewitness accuracy. Studies in this area have generated mixed results; some studies (Clifford & Scott, 1978; Loftus, 1975; Loftus & Burns, 1982; Loftus, Miller, & Burns, 1978; Loftus & Palmer, 1974) show robust effects for eyewitness inaccuracies (e.g., high suggestibility, poor recall), while others show that eyewitness testimony is fairly accurate (Eugenio, Buckhout, Kostas, & Ellison, 1982; Scrivner & Safer, 1988; Turtle & Yuille, 1994; Yuille & Cutshall, 1986). Migueles and Garcia-Bajos (1999) attribute these equivocal findings to "the diversity of research methods used...and the different tests utilized" (p. 257). This is an indication of the need for standardized instruments to be developed to study this phenomenon more uniformly across different laboratories. The purpose of the current investigation was to develop a new assessment tool to measure eyewitness accuracy and suggestibility in college-age adults. Furthermore, the current study examined how suggestibility is related to performance in target-present and target-absent lineups.

Eyewitness Accuracy

Eyewitness accuracy is most commonly studied by examining three different areas. First, the eyewitness's ability to correctly recall the details of the event is measured (free recall). Second, the eyewitness's responses to accurate or inaccurate probe questions are examined. Third, the eyewitness's ability to choose the culprit from a lineup is evaluated. These three components, free recall, responses to probe questions, and lineup performance are the most common ways that researchers study the accuracy of eyewitness memory.

Free Recall

In general, free recall given by an eyewitness immediately or shortly after the witnessed event has taken place is highly accurate (Clifford & Scott, 1978). Clifford and Scott examined the accuracy of free recall versus answers to closed-ended (i.e., yes/no) probe questions after participants had watched a videotape of either a violent event or a non-violent event. They found that statements made during free recall were more accurate than were participants' answers to interrogative questioning. Using these procedures, participants might have been more likely to accept misinformation during the questioning procedure that they normally would not have spontaneously generated. Furthermore, these researchers found that free recall for a violent video was much poorer than the free recall for a nonviolent video. That is, participants recalled more accurate details about the nonviolent witnessed event than the violent event.

This was further demonstrated by Clifford and Hollin (1981) who found again that accuracy of free recall for nonviolent videos is superior to that of violent videos. They also found that when there are several perpetrators, free recall accuracy declines, but only when participants were exposed to violent videos. Accuracy of free recall for nonviolent videos was unaffected by number of perpetrators.

Accuracy of free recall is affected by the encoding process (Eugenio et al., 1982). The accuracy of one's free recall depends upon whether or not one grasped the gist of the situation. Gist is defined as understanding the basic content of the witnessed event or being able to recount the major events that were witnessed. Eugenio and colleagues manipulated a stimulus in which two videos were shown to the participant; one video that showed events unfold in an orderly fashion and one video that was more difficult to understand (i.e., the actors actions were difficult to follow). These researchers found that when shown a video of a crime, individuals who understood the video and were able to keep track of the actors and what they were doing had many more accurate details in their free recall than individuals who did not grasp the gist of the video.

Additionally, Eugenio and colleagues studied hypermnesia, an increase in amount of information recalled over several different time periods and interviews (Scribner & Safer, 1988). Participants who had the correct gist of the video showed more hypermnesia than did those who did not have the correct gist of the video. Furthermore, those who had the correct gist showed superior accuracy of free recall over those who did not have the gist. Thus it appears that when individuals can understand what they are seeing and logically follow events, they can give increasingly accurate accounts of events during multiple free recall sessions.

Yuille and Cutshall (1986) conducted a study in which individuals were interviewed about a live, non-staged crime they had witnessed. This study examined how these findings from the laboratory translate into eyewitness testimony about an actual witnessed crime. In this study, Yuille and Cutshall were able to interview individuals who had witnessed a burglary that resulted in a shooting death. The individuals in the study were interviewed immediately by police officers after the shooting. They were interviewed again (this time by researchers) 4-5

months later. Yuille and Cutshall found that, of the 13 individuals who agreed to the “research interview,” there was little change in the accuracy of their recall over a 5-month period. Individuals were extremely accurate, recalling information about physical attributes with 76% accuracy. Objects were recalled with a rate of 89% accuracy. There were only 3 cases in which accuracy changed in any significant way between the interview sessions. These findings must be interpreted with some caution, as the entire sample included only 13 individuals.

However, Yuille and Cutshall’s findings may not be generalizable to all real world settings, as witnesses might come forward only after a significant delay between when the event occurred and when they speak to a police officer about the event. There is also a significant time lag between when depositions are given and courtroom proceedings take place. Individuals might speak to other eyewitnesses or listen to erroneous television reports that give misinformation about the event. Additionally, lawyers or other individuals who interview them about the event might give them suggestions, having a negative influence on the accuracy of their testimony.

Impact of Misinformation on Open-Ended Questions

Although individuals’ free recall is generally accurate, misinformation can be introduced into their memories. Much interest lies in what happens between when the individual initially witnesses an event and when an eyewitness gives a deposition or testifies in a courtroom. Exposure to misinformation after the individual witnesses an event can alter his or her memory, making future free recall reports erroneous (Clifford & Scott, 1978; Loftus, 1975; Loftus & Burns, 1982; Loftus, Miller, & Burns, 1978; Loftus & Palmer, 1974).

Exposure to misinformation can impinge on the accuracy of an individual’s memory for an event (Loftus & Hoffman, 1989). This could be caused by a number of reasons, particularly

the “weakening of memory traces, or a clouding of memory, or an intrinsic impoverishment of memory (Loftus & Hoffman, p. 101). Loftus and Hoffman discuss 4 reasons why acceptance of misinformation can occur. First, an individual might accept misinformation because a memory of the original event was never fully encoded. This could occur because either the individual was not paying close enough attention to the event or he or she was distracted while the event occurred. Of course, if information is not encoded into memory, it cannot be retrieved at a later time. Second, acceptance of misinformation might occur because he or she might trust the other person’s memory (i.e., the giver of the misinformation) more than they trust their own memory. If the interviewer is a person of authority, the individual might feel that the interviewer must know more about the event than he or she does and consequently will be more likely to agree with whatever the interviewer says. Third, guessing might occur. If an individual feels uncomfortable in a situation and is being pressured to give an answer, they might give an incorrect answer instead of simply saying “I don’t know” or “I don’t remember.” Finally, fourth, the original accurate memory could be intact, but misinformation might have altered the memory during the process of retrieval. The misinformation introduced during the interview might interfere with the retrieval of the memory for the initial event.

Empirical research supports the idea that post-event misinformation can distort memory (e.g., Loftus, 1975; Loftus, Miller & Burns, 1978; Loftus & Palmer, 1974). Manipulation of the contents of a question can alter the ways in which individuals recall information (Loftus & Palmer, 1974). After watching films of vehicular accidents, participants were asked “About how fast were the cars going when they hit each other” (Loftus & Palmer, p. 586). The verb *hit* was then replaced with either *smashed*, *collided*, *bumped*, or *contacted*. Participants estimated that the cars had been going 40.8 miles per hour when the word *smashed* was used. When the word

contacted used, participants estimated that the cars had been going only 31.8 miles per hour. Loftus and Palmer concluded that the way in which the question was phrased influenced the participant's memory for the film. When *smashed* was used, they recalled that the cars had been going very fast, but when *contacted* was used, they did not recall that the cars were traveling at a high rate of speed.

In a follow-up experiment, Loftus and Palmer (1974) found that participants who had been interviewed using the word *smashed* falsely recalled broken glass from the video. Participants who had been interviewed using the word *hit* correctly reported that no broken glass was present in the video.

In a series of experiments, Loftus (1975) further showed how the wording of questions can influence an individual's memory for a witnessed event. These experiments offered support for the construction hypothesis, which states that individuals will try to "visualize or reconstruct" (Loftus, p. 564) parts of the witnessed event when asked questions about it. Loftus found that participants will review the contents of their memories when asked about true events. For example, participants watched a film in which a car speeds through a road sign and causes an accident. Participants viewed either a stop sign or a yield sign. When participants viewed the yield sign, they were asked whether or not they saw a stop sign. Fifty-nine percent of these individuals reported that they had seen a stop sign when in reality it had been a yield sign in the video.

Loftus expected that if the participant would accept that the stop sign was there by constructing a memory of it, participants might also construct memories of events that never happened or objects that were never seen. Loftus (1975, experiment 2) found that this does, in fact, occur. Participants were shown a video containing 8 individuals. Participants were given

inaccurate post-event information, either that there had been 12 individuals in the video or there had been 4 individuals in the video. Individuals in these groups recalled either 8 individuals or 6 individuals, respectively. If nothing else, this experiment supports the idea that the number of perpetrators recalled can vary greatly from one individual to another.

Further support for the construction hypothesis was found in two other experiments by Loftus (1975). These experiments indicated that the wording of questions can influence whether or not participants will falsely recall objects seen in a video. One video showed another vehicle accident and the other showed a vehicle that ran into a man with a baby carriage. In the first experiment, Loftus was able to make individuals falsely recall seeing a barn, by asking “How fast was the car going when it passed the barn” (p. 566). Although there was no barn, 17.3% of individuals reported seeing a barn, while only 2.7% of individuals in the control group reported seeing a barn. In the second experiment, 29% of the participants erroneously reported seeing a school bus after they had been asked, “Did you see the children getting on the school bus?” (p. 567). Only 8% of participants in the control group erroneously reported seeing the school bus.

Suggestibility

Misinformation could also be presented to an individual by asking inaccurate leading questions (see Gudjonsson, 1992). This technique introduces misinformation into the interview discreetly and individuals will often incorporate these pieces of inaccurate information presented by the interviewer into their memories. One’s general tendency to accept misinformation given by an interviewer is referred to as suggestibility (Gudjonsson, 1984).

Individual differences in suggestibility have important real-world applications, particularly in the area of evaluation of eyewitness testimony. Individuals who are highly suggestible may incorporate false information into their memories or succumb to interviewer

pressure. Consequently, innocent individuals have been sentenced to lengthy prison terms based on faulty testimony. The study of individual differences in suggestibility is important because it might be a trait that some individuals have. Although suggestibility can be seen as a state characteristic, there is also evidence that it is a trait characteristic (Gudjonsson, 1992). There is a need to understand why some individuals are suggestible while other individuals are not as suggestible.

Individual differences in suggestibility have been widely studied by Gisli Gudjonsson and colleagues (see Gudjonsson, 1992; 2003). Gudjonsson studies interrogative suggestibility, which is defined as the acceptance of misleading information given during a questioning procedure (Gudjonsson & Clark, 1986). Gudjonsson (1987a) differentiates interrogative suggestibility from more traditional types of suggestibility, such as hypnotic suggestibility or idio-motor suggestibility. These types of suggestibility are often associated with either attitude changes (e.g., the person is led to believe something contradictory to their actual beliefs) or a false experience of some sort of sensory stimulation (e.g., the person is led to feel a non-existent heat source). Interrogative suggestibility is more associated with how an individual's verbal recall for an experienced event might be changed by the misinformation provided during a questioning procedure.

Two types of approaches to the study of interrogative suggestibility include the individual differences approach and the experimental approach. Gudjonsson is associated with the individual differences approach, while other researchers, including Loftus and her colleagues (e.g., Loftus, Miller, and Burns, 1978) use the experimental approach. The individual differences approach examines how individuals use different coping strategies in response to uncertainty. Some individuals will give in to the misinformation to cope with the uncertainty,

while others feel comfortable reporting that they “don’t know” or “don’t remember” the details of a given situation. The experimental approach examines how different types of interviewing techniques can influence eyewitness memory. These approaches have much to offer each other and should be considered complimentary rather than opposing views.

According to Gudjonsson and Clark (1986), there are five elements associated with interrogative suggestibility, which are “the nature of the social interaction, a questioning procedure, a suggestive stimulus question, some form of acceptance of the stimulus message, and a behavioral response” (p. 84). There must be a social interaction that includes an interviewer and an interviewee. The suggestive stimulus question would be an inaccurate leading question. This question would be asked by the interviewer to discern whether or not the interviewee will accept the misinformation given in the inaccurate leading question. Finally, the inaccurate leading question must influence an individual’s behavioral response during an interrogation (i.e., he or she will modify his or her story in response to the misinformation).

The elements of interrogative suggestibility describe the features that are necessary in order to measure interrogative suggestibility. In addition to the five elements of interrogative suggestibility, Gudjonsson & Clark (1986) define three components of interrogative suggestibility. The components of interrogative suggestibility describe the factors that can bring about the behavioral response in the individual (i.e., high levels of suggestibility). These components are a) uncertainty, b) expectations of success, and c) interpersonal trust. These factors must be present in order for a suggestive response to occur.

First, the individual must be faced with uncertainty; he or she does not know the correct answer to a question, so there is some anxiety present over giving the proper or desired answer. This is in contrast to compliance, which occurs when the individual knows the correct answer,

but gives a wrong answer to please the interviewer. Second, the individual must feel trust in the interviewer. That is, the individual must feel that the interviewer is not trying to be tricky or deceitful. Trust allows the interviewee to feel comfortable accepting information from the interviewer. Finally, the individual must feel some degree of expectancy for success. They want to give accurate testimony and want to be perceived as having the correct answers (Gudjonsson & Clark, 1986). Taken together, these three components are necessary conditions for acceptance of the suggestion to occur.

Assessment of Suggestibility

The Gudjonsson Suggestibility Scales (GSS). Up until the early 1980's, there was no measure to assess an individual's suggestibility. Gudjonsson (1984) reports that previous studies of suggestibility had been closely associated with hypnotic suggestibility, rather than interrogative suggestibility. Addressing this need, Gudjonsson (1984) developed a scale to measure interrogative suggestibility in adults. The Gudjonsson Suggestibility Scale (GSS1) involves listening to a short narrative about a crime, and subsequent questioning about what the individual heard in the narrative. The interview contains 20 questions, five accurate and 15 misleading. This interview is administered 40-50 minutes after hearing the narrative. After initially asking the 20 questions, the interviewer provides negative feedback (e.g., "you missed several of the questions), and the questions are asked again. The reason for providing the negative feedback and asking the questions again is to measure an individual's tendency to change his or her answers.

Gudjonsson (1987) developed a parallel form of the GSS1, called the GSS2. The GSS2 is the same as the GSS1 except that the GSS2 does not have a criminal element to the narrative that individuals listen to. Gudjonsson has found that individuals perform the same on the GSS2

as they do on the GSS1. This has led him to believe that “interrogative suggestibility can be reliably measured employing varied narrative content” (Gudjonsson, p. 219). Therefore, the GSS1 and GSS2 can be used rather interchangeably.

Yield and shift. Gudjonsson (1984) viewed interrogative suggestibility as consisting of two factors, Yield and Shift. Yield refers to the tendency of an individual to accept misinformation presented during a questioning procedure. In Gudjonsson’s scale, there are 15 questions asked that contain misleading information. When an individual responds affirmatively to one of these 15 misleading questions, this is scored as a yield. The second component, Shift, refers to the change in responses given after the interviewer provides the participant with negative feedback. An individual can shift from either a correct answer to an incorrect answer or from an incorrect answer to a correct answer. The sum of the Yield and Shift gives the total suggestibility score.

As a construct, Yield reflects a cognitive component. An individual might acquiesce to misinformation because he or she cannot recall the events in which he or she witnessed. This would reflect a memory deficit, as the individual accepts misinformation because he or she cannot correctly recall the correct information. Shift refers to a social component. After the individual has been asked all of the questions, he or she is told that many of the questions were answered incorrectly and that they must try to do better next time. When an individual shifts his or her answer, this is largely because of the social pressure to do so (Gudjonsson, 2002??)

Yield can be further divided into Yield 1 and Yield 2. Yield 1 refers to the number of items of misinformation the individual accepts before the negative feedback is presented (Gudjonsson, 1992). Yield 2 refers to the number of items of misinformation the individual accepts after the negative feedback is given. Therefore, Yield 2 is a reflection of how much

misinformation the individual has accepted as evidenced by his or her acquiescence to the misleading questions when they are asked the second time. According to Gudjonsson, Yield 2 is a better indicator of suggestibility, as it measures individuals' responses after they have been placed under "interrogative pressure" (p. 135). However, Yield 1 is used in conjunction with Shift to score total suggestibility.

The GSS1 and GSS 2 both have been used in a variety of settings and age groups, including normal controls (Baxter & Boon, 2000; Gudjonsson, 1983; Gudjonsson, 1986; Gudjonsson, 1988; Gudjonsson, 1989; Gudjonsson, 1991), forensic patients (Clare & Gudjonsson, 1993; Gudjonsson, 1990; Gudjonsson, 1991; Gudjonsson, Rutter, & Clare, 1995; Gudjonsson & Sigurdsson, 1996; Sigurdsson & Gudjonsson, 1996), and adolescents (Gudjonsson & Singh, 1984; Richardson, Gudjonsson, & Kelly, 1995; Sigurdsson & Gudjonsson, 1996). The findings of each of these studies are discussed below.

Normal controls. In normal controls, suggestibility was found to be related to intelligence, personality, and memory abilities (Gudjonsson, 1983). Specifically, total suggestibility was positively related to neuroticism (as measured by the Eysenck Personality Questionnaire). However, when analyzed separately Yield and Shift subscales alone did not correlate with any of the personality dimensions measured (i.e., psychoticism, neuroticism, extraversion, and social desirability). This is expected, as those who are high on neuroticism often want to present themselves in a favorable light. To further support this finding, Gudjonsson (1988) found that suggestibility relates negatively to assertiveness (as measured by the Rathus Assertiveness Schedule) and positively with anxiety (as measured by the Spielberger State Anxiety Inventory).

Total suggestibility, Yield 1, and Shift are all positively related to acquiescence, which is one's general tendency to respond with an assent to questions asked in an interview setting (Gudjonsson, 1986). This effect is especially strong on the Shift subscale, as it is a response to pressure from negative feedback. Individuals may feel pressure to answer affirmatively to questions asked by the interviewer once they have told that their previous responses were wrong. This is in agreement with the findings discussed previously concerning the relationship between suggestibility and anxiety. Once the negative feedback is administered, individuals might begin to feel anxious, causing them to answer affirmatively to the questions asked by the interviewer.

The administration of the negative feedback can also affect the extent to which individuals will yield or shift. Baxter and Boon (2000) found that when the negative feedback is delivered in a stern manner, individuals will have higher Shift scores and higher Yield 2 scores than individuals who experienced the negative feedback in a friendly or firm manner. Therefore, when individuals feel that an interviewer is being harsh or demanding, the individual will be more likely to submit to the negative feedback and either accept the misinformation or change his or her answers.

Forensic populations. Suggestibility levels are similar in forensic populations as they are in normal control populations (Richardson et al., 1995). Richardson and colleagues found that among adolescent boys, those who are delinquent have similar Yield scores to those who are nondelinquent. Furthermore, in forensic populations, adolescents are more suggestible and are more likely to shift than are adults. Richardson and colleagues (1995) conclude that adolescents who are delinquent are particularly susceptible to suggestion and that care should be taken in order to not put excessive pressure on them during interviews. This effect has not been shown in an adult population, and it warrants further investigation.

In adult forensic populations, the GSS has been used extensively. Gudjonsson and Sigurdsson (1996) examined correlates of suggestibility among Icelandic prison inmates. They found that the prison inmates who were of lower intelligence had higher Shift scores than those who were more highly intelligent. Gudjonsson and Clare (1995) performed the same experiment on normal controls and did not find a relationship between intelligence and Shift. Comparing the two studies, the prison inmates had lower scores on Yield 1, Yield 2, and Total Suggestibility (mean scores = 4.4, 6.7, and 9.4, respectively) in comparison to the Yield 1, Yield 2, and Total Suggestibility scores of the normal controls used in the Gudjonsson and Clare study (mean scores = 8.0, 8.2, and 12.7, respectively).

The Video Suggestibility Scale for Children. The Video Suggestibility Scale for Children (VSSC) was created by Scullin & Ceci (2001). This scale was modeled after the Gudjonsson Suggestibility Scales (Gudjonsson, 1984; 1987). The VSSC consists of a 5-minute video of a child's birthday party and a subsequent suggestive interview about the video. The interview contains a free recall component and a list of 18 probe questions. Fourteen of these questions contain misleading information, such as the suggestion that there was a clown at the party when in reality there was not. If the child affirms the suggestion offered in a question, this is scored as a yield (Scullin & Ceci, 2001). Upon completion of the first nine questions, the child is told that he or she made several mistakes and that they were going to have to be asked the questions again to "see if they can do better." If a child changes his or her answer to a question after this negative feedback is given, this is scored as a shift. Total suggestibility is defined by the sum of both the yield and the shift responses. Scullin, Kanaya, and Ceci (2002) found that the VSSC predicted suggestibility for children over the age of 4 ½ (54 months) but not for children under this age.

Eyewitness Identification

Memory for physical attributes. Individuals' ability to recall information about physical attributes of another individual is typically quite accurate (Yuille & Cutshall, 1986). Recalling such details as height, weight, hair color, and race seems to be a fairly easy task for most individuals. According to Tollestrup, Turtle, and Yuille (1994), individuals can describe a perpetrator's age, weight, and height with such accuracy that they only missed by one or two years within actual age, a few pounds within actual weight, and only a few inches within actual height. Ebbesen and Rienick (1998) reported that accuracy of physical descriptions of target individuals did not decline over a period of 4 weeks.

Lineup Performance

Target present vs. target absent. The target-present lineup is one in which the suspect is one in which a suspect is placed in a lineup along with several other fillers (i.e., people who are not suspects). The target-absent lineup is one in which a known innocent individual is placed in a lineup along with several other fillers (Wells & Olson, 2003). Target-absent lineups are useful in determining the accuracy of a witness; if he or she selects a target from a target-absent lineup, it can be concluded that he or she does not have a strong memory for what that person looked like. Thus, the police will no longer use the witness.

Wells and colleagues (1998) proposed several recommendations for conducting lineups. The first recommendation is that the lineup should be double blind, meaning that the officer in administering the lineup should not know who the suspect is. This will prevent the officer from giving any feedback to the individual, confirmatory or otherwise. Second, Wells and colleagues say that the individual should be given instructions that are nonbiased. In other words, the individual should be warned each time that the suspect may or may not be present in the lineup. Third, the distracters in the lineup should match the description of the culprit, as provided by the

individual. If this condition is not met, it may result in a biased lineup, making it easy for the individual to choose the target because he or she appears so different from the others. Finally, the last recommendation is that confidence ratings should be taken immediately after the individual selects a person from the lineup. Kassir (1998) recommends that a fifth rule be added, which is that the lineup procedures should be videotaped. According to Kassir, this will provide an objective assessment of the fairness of the lineup.

Signal detection theory and identification. Some studies have examined the link between signal detection theory and facial identification. Signal detection theory deals with those factors that are concerned with making sensory judgments (Prkachin, 2003). People have different thresholds for which they are willing to say that they detect a stimulus. When faced with a stimulus such as a series of faces, the individual has certain criteria that must be present in order for him or her to classify the face as one that has been seen before or one that has not been seen before.

Signal detection analysis works in the following way: When the stimulus is present and the participant says that he or she detects it, this is called a hit. When the stimulus is present and the participant says that he or she does not detect it, this is called a miss. When the stimulus is not present and the participant says he does not detect it, this is called a correct rejection. Finally, when the stimulus is not present and the participant says that he or she detects it, this is called a false alarm (Prkachin, 2003).

Park, Lee, and Lee (1996) applied signal detection theory to facial identification ability. They showed participants a video of a burglary and later asked them to select the culprit from a series of photos. In these photos, the angle in which the photo was taken and the attire of the culprit were changed. Park and colleagues found that when the individuals were more

perceptually sensitive (i.e., attuned to changes in attire), they used a more sensitive judgment criterion. Therefore, they were able to select the culprit from the photos regardless of angle or attire because they were sensitive to small changes and were not fooled by slight differences. Park and colleagues concluded that when individuals are more observant of their surroundings, they will be more likely to make a large number of hits in comparison to those who are not as observant of their surroundings.

Limitations of the Eyewitness Identification Literature

Considering the literature reviewed, there is a clear need for a new measure of suggestibility. The shortcomings of The GSS1 and the GSS2 include 1) auditory presentation rather than both auditory and visual presentation, 2) a short time interval between the event and the interview, and 3) the small number of accurate leading questions on the scale.

The procedures involved in the GSS1 and the GSS2 entail listening to a short story, with either a forensic content or a non-forensic content (GSS1 and 2, respectively). This is the only means of presentation; this does not permit the researcher to explore crucial area of eyewitness accuracy, namely identification of perpetrators. Furthermore, the applicability of this to a real world situation is limited, as individuals rarely only overhear a crime. Individuals who witness a crime both see and hear the events, and are subsequently interviewed about both what they saw and what they heard. Thus, a scale that incorporates both a sensory and an auditory component is needed.

The Video Suggestibility Scale for Adults (VSSA) will involve both a visual and an auditory component. Individuals will watch a video, depicting perpetrators who commit two crimes, one violent in nature and one that is nonviolent. Some of the actors in the videos will be clearly displayed so that the individuals have ample time to form a memory of their features.

However, two individuals in the video are not clearly displayed and it might prove to be a challenge to correctly identify them. Thus, it will not be too difficult, nor will it be too easy to select the perpetrators from the lineups.

Second, the GSS1 and the GSS2 are limited because of the short interval between when the story is read to the participant and when they are interviewed about what they heard. This is somewhat true to what happens in real life situations, as a policeman might arrive on the scene and interview the witnesses right away. However, there will often be a delay of several days, weeks, or months between when the crime happened and when the individual is interviewed about it. The VSSA examined differences between individuals who were immediately interviewed after viewing the videos and individuals who were interviewed after a 5-8 day delay. This is more true to life and it also allows for some forgetting to occur. Because accuracy of memory declines over time (Kassin, Ellsworth, & Smith, 1989) it is important to allow time for this process to occur. Interviewing the participants too soon after watching the video might result in an unrealistically high degree of accuracy.

Third, the GSS1 and the GSS2 have not incorporated many accurate questions into the interview. This might allow the individual to more easily guess the true nature of the scale. In other words, the individual might feel that he or she is being tricked. More accurate questions on the scale might help to disguise the true nature of the scale. The VSSA will incorporate more accurate questions into the interview. The addition of more accurate leading questions might also elicit more shifting responses from an accurate “yes” response to an inaccurate “no” response. This will demonstrate the effectiveness of the negative feedback. If the negative feedback is successful, the individual might experience doubt over his or her first correct answer and choose to change his or her answer to an inaccurate response.

The Video Suggestibility Scale for Children (Scullin & Ceci, 2001; Scullin et al., 2002) overcame these limitations and has shown to be a reliable and valid tool for measuring children's suggestibility (with the exception of an even number of accurate and inaccurate questions). The Video Suggestibility Scale for Adults is modeled after this scale, and goes further by adding the identification component. The scale is expected to be a useful tool for measuring suggestibility in adults.

Another limitation in the eyewitness identification literature is that there have been no studies that examine the relationship between suggestibility and an individual's ability to correctly identify a culprit from a lineup. This is important to study as it may clarify how individuals who are highly suggestible perform on lineups. If it is found that suggestible people have many false positives on target-absent lineups, this does not indicate that their testimony cannot be trusted. Rather, it identifies individuals high in suggestibility as a unique group who might need special instructions when attempting to identify a witness in a real-world forensic setting.

One additional advantage it allowed the researcher to control variables such as amount of time that has passed between witnessing the event and being questioned about it. Additionally, variables such as physical facial changes and amount of light can also be controlled for.

Statement of the Problem

The current study proposes the development of a new suggestibility scale for adults, the Video Suggestibility Scale for Adults (VSSA). This scale was modeled after the Video Suggestibility Scale for Children (Scullin & Ceci, 2001; Scullin et al., 2002), which has shown to be a reliable and valid tool for measuring interrogative suggestibility in preschool-age children. The development of the VSSA has arisen from the need to measure suggestibility in a

standardized way, while using more real-world procedures than have been used in past studies. The current study will also try to establish a relationship between suggestibility, as measured by the VSSA and lineup performance.

Participants watched the videos of the crimes and were subsequently interviewed about what they had seen. This interview consisted of a free recall portion (e.g., tell me everything you remember about the carjacking and the burglary) and a series of probe questions were asked. Participants then viewed 12 lineups, four of them were target-present (i.e., contained a character from the video) and eight of them were target-absent lineups.

Research Questions and Hypotheses

Research question #1. Is the VSSA useful tool for studying suggestibility in college aged adults?

Hypothesis 1: Adults will respond to the leading questions in the interview similarly to how children did with the VSSC and how adults responded to the GSS. Adults will yield to the suggestions made by the interviewer. Scullin and Ceci (2001) found that children often yielded to leading questions, while Gudjonsson (1984) has found that adults also will yield in response to leading questions.

Hypothesis 2. Adults will shift their answers in response to negative feedback from the interviewer. Scullin and Ceci (2001) found that children often shift their answers in response to negative feedback, while Gudjonsson (1984) has found that adults also will shift in response to negative feedback.

Hypothesis 3. The higher number of accurate statements in free recall will be associated with lower total levels of suggestibility.

Research question #2. Will there be differences between those individuals who are interviewed immediately and those who are interviewed after a 5-8 day delay?

Hypothesis 4. Individuals in the immediate condition will have a higher total number of statements generated in their free recall in comparison to those in the delayed condition.

Hypothesis 5: Individuals in the immediate condition will have a higher total number of accurate statements generated in their free recall in comparison to those in the delayed condition.

Hypothesis 6: Individuals in the immediate condition will have a lower total number of inaccurate statements generated in their free recall in comparison to those in the delayed condition.

Hypothesis 7: Individuals in the immediate condition will have lower Yield, Shift, and Total Suggestibility scores in comparison to those in the delayed condition.

Research question #3. How does lineup performance relate to suggestibility?

Hypothesis 8. Highly suggestible individuals will be more likely to falsely identify a witness from a target-absent lineup. Participants who score high on suggestibility will show poor performance on the target-absent lineups. This hypothesis is exploratory.

Hypothesis 9. Individuals who score high on Yield 2 will be more likely to identify a witness from a target-absent lineup. Yield 2 is a strong indicator of suggestibility and it is also an indication that the individual has accepted misinformation (Gudjonsson, 1992). Therefore, it follows that an individual who scores highly on Yield 2 will also be highly suggestible and will be likely to falsely identify someone from a target-absent lineup.

Hypothesis 10. There will be no difference in performance between highly suggestible individuals and non-suggestible individuals on the target-present lineups. This hypothesis is exploratory.

Method

Participants

A power analysis, using Sample Power, indicated that with a sample size of 168 and an alpha set at .05 the study would have power of .80. This indicates that this sample size is large enough to detect medium size effects 80% of the time. One hundred eighty-one undergraduates were recruited from psychology classes to participate in this study.

Attrition. Due to participants not keeping appointments and problems with missing data, the final sample size was 160 (n immediate condition = 86, n delayed condition = 74). Twenty one participants completed the first portion of the interview, but did not return for the second part of the interview. These individuals differed from those individuals who kept their appointments in two ways; individuals who did not keep their appointments were significantly older than those who did keep their appointments, $F(2, 178) = 4.817, p > .05, \eta^2 = .051$ (see Table 1). Follow-up t -tests indicated that participants in the immediate group and the delayed group both differed significantly from those in the never returned group, $t(105) = 1.46, p = .008$ and $t(93) = 2.58, p = .04$, respectively. These groups also differed in regard to class rank, with the groups having significantly different numbers of freshman, sophomores, juniors, and seniors, $X^2(6, N = 180) = 16.78, p = .01$ (see Table 1). The groups did not differ significantly in regard to GPA, race, or gender.

Among those participants whose data could be used, the participants' ages ranged from 18 to 26 years old, with a mean of 19.6 ($SD = 1.5$). The sample was overwhelmingly female (83%) and Caucasian (74.9%). Asian Americans comprised 5.6% of the sample, followed by African Americans (3.8%), Other (3.8%), and Native Americans (0.6%). The sample contained

mainly freshmen (34.4%), but sophomores, juniors, and seniors were also represented (28.1%, 21.9%, and 15.6%, respectively).

Measures

Suggestibility. The Video Suggestibility Scale for Adults was used to measure suggestibility. This scale utilizes videos of crimes that were staged by the researchers. The first crime features two men who steal a woman's car after hitting her on the head and knocking her down. The second crime features three men who burglarize a neighbor's house. The video was shot using a hand-held digital camera. The actors in the video were the author's friends from her hometown. These individuals were selected to avoid the problem of the participants recognizing someone from campus.

In the first part of the video, two men are hanging out in a parking lot, discussing what kind of trouble they want to get in to that night. They decide to steal a car. As they are looking for a car to steal, they notice a good-looking woman leaving a drug store. They notice that she is driving a nice car so they decide to steal her car. As she approaches her car, the men are yelling and taunting her. They attack her, hit her on the head with a bat, and take her purse and keys. They drive off. In the next scene, the carjackers are seen counting the money from her purse. They abandon the car and run into the woods.

In the second part of the video, three men are sitting in a room discussing the possibility of burglarizing a neighbor's house. The man holding the camera talks them into robbing a certain neighbor's house. In the next scene, the men are seen breaking into the house. After they get in, they survey the house and discuss what they are going to take. One of the men hears a noise and looks out the window, fearing that the homeowners have come home. Hearing nothing, he resumes the burglary. A few moments later, however, a little girl walks into the

house, sees the burglars, and yells for her mother. The burglars leave quickly, taking with them video games, jewelry, and a bass guitar. In the final scene, the burglars returned home and were laughing about their crime, very proud of what they have done.

The videos were shot without using a script in an attempt to maintain the perception that these videos were not staged. Furthermore, they were intentionally made to look rough, as if these were really criminals who were videotaping their crimes. As a result, the videos appear to be unsteady and unclear at times.

Sign-up procedure. A sign-up sheet was hung on the Participant Recruitment Board in the psychology department. Participants signed up in pairs, so that two participants would show up for each appointment time. When the participants arrived in the lab, both of them would watch the video and both of them would complete the Multidimensional Personality Questionnaire. Upon completion of this, one participant was randomly chosen to be interviewed about the video immediately and one participant was assigned to the delayed interview condition. This participant then made an appointment to finish the interview after a 5-8 day delay. Twenty-one participants signed up for this delayed interview condition and did not return for the appointment (discussed earlier).

Interviews. The interview about the video consists of a free recall component followed by a series of probe questions. First, the interviewer asks the participant to think back to the video about the carjacking. Then the interviewer instructs the participant to tell him or her everything that he or she remembers about the video, even things that aren't important. After the participant has finished recalling the events of the carjacking, he or she is asked about the burglary. After the participant has recalled the events of the burglary, a series of 92 probe questions are asked. The questions in the interview are either accurate leading (i.e., they contain

information that actually happened in the video) or inaccurate leading questions (i.e., they contain information about events that did not happen in the video).

This scale, developed by Bonner and Scullin, was pilot tested with 129 undergraduates. The results of the pilot study indicate that college aged students are responding in ways similar to how preschoolers responded to the VSSC and how adults responded to the GSS. The pilot scale was composed of 44 accurate questions and 67 inaccurate questions.

For the accurate questions, the scale yielded a coefficient alpha of .69. For the inaccurate questions, the scale yielded a coefficient alpha of .57. The mean score for the accurate questions was 29.8 ($SD = 4.8$) and the range was 9 – 40, out of a possible 44 questions. The mean score for the inaccurate questions (Yield 1) was 12.2 ($SD = 5.4$) and the range was 2 - 31, out of a possible 67 questions. The mean number of Shifts was 17.6 ($SD = 7.46$) and the range was from 2-38, out of a possible 111 questions.

The coefficient alpha for the scale was rather low, so an item analysis was conducted to determine which questions should be deleted from the scale in order to increase reliability. Nineteen questions were removed from the initial scale, leaving 92 total questions, 52 accurate and 40 inaccurate.

Interviewer training. Over the course of data collection, nine female interviewers and two male interviewers were trained. The interviewers did not see the video, as doing so might have contaminated the interview; had the interviewers seen the video, they might have inadvertently led participants to a correct answer or led them to select the correct culprit from the lineup.

The interviewers were given the interview packet and told to engage in practice interviews until the questioning procedure felt natural. The author and other graduate students

from the lab would act as subjects and assess the interviewer's skill. The interviewers were assessed by the following criteria: a) not reading from the paper, b) maintaining eye contact with the participant, and c) maintaining a stern tone when delivering feedback. The author was present for the interviews' first several interviews to ensure that the procedures were being followed.

Overall, 27 participants were interviewed by a male interviewer (11 males and 16 females). One hundred thirty-three participants were interviewed by one of the 9 female interviewers (48 males and 85 females). A chi-square test indicated that these differences were not significant; males did not significantly interview one gender more than the other nor did females significantly interview one gender more than the other.

This same result was found when looking at each condition. In the immediate group, 18 participants were interviewed by a male interviewer (8 males and 10 females). Sixty-eight participants in the immediate group were interviewed by a female (22 males and 46 females). In the delayed group, 9 participants were interviewed by a male interviewer (3 males and 6 females). Sixty-five participants in the immediate group were interviewed by a female (26 males and 39 females).

Free recall coding. Three independent coders who were not conducting interviews coded the free recall portions of the transcript. In line with procedures used by Gudjonsson (1984), the portion of the interview in which the participant is asked to "tell everything you remember about the burglary and the carjacking" was coded for total number of statements generated, total number of accurate statements generated and total number of inaccurate statements generated.

The coders watched the video and created a script based on the events that took place. The coders developed a list of 78 salient points that occurred in the video. The coders read each

transcript and broke each statement into meaningful parts. Each unit was then coded as being accurate, inaccurate, or other. The coding scheme used was similar to the coding scheme used with children by Scullin et al. (2002). Scullin and colleagues examined accuracy or inaccuracy at the “utterance level” (p. 237). According to these researchers, an utterance “contains a verb and is bounded by a pause” (p. 237). For example, the statement “two guys were sitting in a parking lot and admiring a black car” would be divided into two utterances (“two guys were sitting in a parking lot” and “admiring a black car.”) These two utterances would be then assigned three points for accuracy; one point for “2 guys sitting in a parking lot”, one point for “admiring a car”, and one point for identifying the car as “black.” Inter-rater reliability yielded an intraclass correlation of .94 for the total number of statements, an intraclass correlation of .89 for the total number of accurate statements and an intraclass correlation of .83 for total number of inaccurate statements.

Lineups. Participants were shown various pictures of individuals in simultaneous lineups. They were shown 6 pictures of individuals and asked to select who they remembered seeing in the video. There were target-present and target-absent lineups for the main characters in the video. In the target-present lineups, there were five foils and the picture of the target individual. In the target-absent lineup, there were six pictures of individuals who were not shown in the video. The foils for these lineups were chosen using the appropriate foils technique. For each target-absent lineup, individuals were chosen who resembled each of the characters from the video. For example, a target-present lineup with the carjacker contains a picture of the carjacker and five other individuals who resembled the carjacker. The corresponding target-absent lineup contained six new individuals who all resembled the carjacker. The same procedures were used for the two burglars and the victim of the carjacking.

Procedure

Participants were brought into the laboratory to watch a video featuring the carjacking and the house burglary. The video is approximately seven minutes long. After watching the video, participants were randomly assigned to be either immediately interviewed or interviewed after a 5-8 day delay.

Immediate condition. After watching the video, participants engaged in a distractor task, which was the completion of the Multidimensional Personality Questionnaire (MPQ). This 155-item scale was used as a distractor to allow for some memory decay to occur. A similar procedure was used by Gudjonsson (1984; 1987). Gudjonsson used a distractor task so that the participants were not interviewed immediately after hearing the story. This was an attempt to allow for some memory decay to occur, and Gudjonsson found that a 20 minute delay was sufficient to allow for this delay to occur. The MPQ takes approximately 30 minutes to complete. This measure was chosen because past studies have shown that suggestibility does correlate with some measures of personality; however, this was not a specific aim of the current study. Upon completion of the MPQ, the participants were administered the following questionnaires: a demographic questionnaire (see Appendix A), the interview component of the Video Suggestibility Scale for Adults (see Appendix B), and the lineups (see Appendix C).

After the administration of the VSSA interview, participants were shown 12 simultaneously-presented lineups. Four of these lineups were target-present and 8 of the lineups were target-absent. The target-present lineups contained 5 foils and the target. The participants were instructed that he or she should examine each face and tell the interviewer whether or not they recognized anyone from the video.

The instructions given to the participant were very general. The interviewer presented the participant with the various lineups and asked if he or she recognized anyone from the videos. The interviewer also cautioned the participant each time that the characters from the video may or may not be in the lineup. No further instructions were given and the interviewer did not give any type of feedback about the correctness or incorrectness of the participant's decision.

Participants were given as long as needed to make their decision. Generally, a decision was made in less than 1 minute. As an incentive for participation, participants were given extra credit in their psychology classes in exchange for their participation.

Results

Preliminary Analyses

The subscales of the VSSA were highly correlated with each other. Yield 1, Yield 2, Shift, and Total Suggestibility were all strongly and positively related to each other. Individuals who were scored highly on Total Suggestibility also scored high on Yield 1, Yield 2, and Shift. These correlations were rather high, ranging from .488 to .733 (see Table 3).

The order of presentation of videos was counterbalanced so that half of the participants saw the burglary first and half of the participants saw the carjacking first. A multivariate analysis of variance (MANOVA) revealed that video order did not influence Yield 1, Yield 2, Shift, or Total Suggestibility scores $F(4, 155) = .874, p > .05, \eta^2 = .022$.

Different interviewers were used throughout data collection. A MANOVA was performed to examine whether or not the gender of interviewer had an influence on Yield 1, Yield 2, Shift or Total Suggestibility scores. The MANOVA revealed that sex of interviewer did

have an influence over these dependent variables, $F(4, 155) = 4.74, p < .001, \eta^2 = .109$. Follow up univariate tests were then conducted to further examine this.

There were no differences between male interviewer and female interviewer on Yield 1 scores $F(1, 158) = 1.733, p > .05, \eta^2 = .011$. With a male interviewer, participants had higher Yield 2 scores than with a female interviewer, $F(1, 158) = 4.97, p < .05, \eta^2 = .031$ (see Table 4). With a male interviewer, individuals had higher Shift scores than with a female interviewer, $F(1, 163) = 7.32, p < .05$ (see Table 4). With a male interviewer, individuals had a higher total suggestibility score than with a female interviewer, $F(1, 163) = 3.68, p < .05$ (see Table 4). The interaction between sex of interviewer and sex of participant was not significant for any of the dependent variables, $F(4, 153) = 1.61, p > .05, \eta^2 = .040$ (Yield 1, Yield 2, Shift, or Total Suggestibility). No main effects for gender of participant were found for Yield 1, Yield 2, Shift, or Total Suggestibility, $F(4, 155) = .970, p > .05, \eta^2 = .024$.

There were small differences found between the participants in the immediate group and participants in the delayed group. Participants in the immediate group were significantly older than participants in the delayed group, $F(1, 158) = 6.24, p < .05, \eta^2 = .038$. The mean age for the immediate group was 19.9 (SD = 1.5) and the mean age for the delayed group was 19.3 (SD = 1.4). Participants in the delayed group, because they were younger, were also more likely to be freshman and sophomores, $\chi^2(3, N = 160) = 8.22, p < .04$. Participants in the immediate condition were mostly juniors and seniors (see Table 1).

Research question #1. Is the VSSA useful tool for studying suggestibility in college aged adults?

Hypotheses 1 and 2 were related to the scale characteristics. These hypotheses stated that Yield 1 and Shift will work similarly in this scale as they did with the VSSC and with the

Gudjonsson Suggestibility Scales. The reliabilities for the scale improved after the deletion of 19 items from the scale used in the pilot sample. For Yield 1, the coefficient alpha was .77. For Yield 2, the coefficient alpha was .85. For Shift, the coefficient alpha was .83.

The mean score for the Accurate Leading Questions was 30.23 ($SD = 4.33$). The scores ranged from 19 – 40 out of a possible 42. The mean score for the Inaccurate Leading Questions (Yield 1) subscale was 6.25 ($SD = 3.65$). The scores ranged from 0 – 21 out of a possible score of 50. The mean score for the Yield 2 subscale was 8.31 ($SD = 5.47$). The scores ranged from 0 to 36 out of a possible score of 50. The mean score for the Shift subscale was 11.39 ($SD = 8.35$). The scores ranged from 0 – 43 out of a possible score of 92. The mean score for Total Suggestibility was 17.64 ($SD = 10.63$). The scores ranged from 0 – 64 out of a possible score of 129.

A principle components factor analysis was performed and a scree test was used to determine the number of factors that emerged. This analysis revealed that the hypothesis that the scale is comprised of two factors was not supported. Forty-eight components with eigenvalues over 1 emerged. Consequently, two factors were rotated using varimax rotation. The varimax rotation procedure indicated that two factors only accounted for 11.9% of the variance. The first factor, Yield 1, accounted for 7.71% of item variance and the second factor, Shift, accounted for 4.2% of the variance. However, the items were not “hanging together” very well. Only 5 of the yield items and 6 of the shift items had a structure coefficient greater than .4.

Hypothesis 3. The higher number of accurate statements in free recall will be associated with lower total levels of suggestibility. This was tested using correlations. The hypothesis was not supported. Total number of accurate statements was not related to total suggestibility, $r = -$

.145, $p > .05$. However, number of accurate statements in free recall was associated negatively related to Yield 1 score, $r = -.183$, $p < .05$.

This was also tested using ANOVA and the proportions of accurate statements made in the free recall. The ANOVA revealed that individuals who were more suggestible generated a significantly lower proportion of accurate statements than those who were less suggestible, $F(1, 145) = 7.88$, $p = .006$, $\eta^2 = .052$. For participants who were more suggestible, their free recall narratives contained 88% accurate statements ($SD = 8\%$). For participants who were less suggestible, their free recall narratives contained 93% accurate statements ($SD = 6\%$).

Other free recall components were analyzed with regard to the VSSA subscales. Total number of inaccurate statements recalled was positively related to Yield 2, Shift, and Total Suggestibility (see Table 3). Participants who generated a high number of inaccurate statements also scored high on Yield 2, Shift, and Total Suggestibility.

Research question #2. Will there be differences between those individuals who are interviewed immediately and those who are interviewed after a 5-8 day delay?

Hypothesis 4 stated that individuals in the immediate condition will have a higher total number of statements generated in their free recall in comparison to those in the delayed condition. This was tested using a one-way ANOVA with type of interview (immediate vs. delayed) as the independent variable and total number of statements as the dependent variable. This hypothesis was supported. Individuals in the immediate condition did generate a higher total number of statements than individuals in the delayed condition $F(1, 153) = 17.02$, $p < .01$, $\eta^2 = .102$ (see Table 5).

Hypothesis 5: Individuals in the immediate condition will have a higher total number of accurate statements generated in their free recall in comparison to those in the delayed condition.

This was tested using a one-way ANOVA with type of interview (immediate vs. delayed) as the independent variable and total number of accurate statements as the dependent variable. This hypothesis was supported. Individuals in the immediate condition did generate a higher number of accurate statements than the delayed condition, $F(1, 153) = 18.30, p < .05, \eta^2 = .109$ (see Table 5).

Hypothesis 6: Individuals in the delayed condition will have a higher total number of inaccurate statements generated in their free recall in comparison to those in the immediate condition. This was tested using a one-way ANOVA with type of interview (immediate vs. delayed) as the independent variable and total number of inaccurate statements as the dependent variable. This hypothesis was not supported. Individuals in the delayed condition did not generate a higher number of inaccurate statements than the delayed condition, $F(1, 154) = .061, p > .05, \eta^2 = .000$ (see Table 5).

Hypothesis 7: Individuals in the immediate condition will have lower Yield 1, Yield 2, Shift, and Total Suggestibility scores in comparison to those in the delayed condition. A MANOVA was conducted with type of interview (immediate vs. delayed) as the independent variable and Yield 1, Shift, and Total Suggestibility as the dependent variables. The MANOVA revealed that type of interview influenced Yield 1, Shift, or Total Suggestibility, $F(4, 155) = 3.78, p < .05, \eta^2 = .089$.

The one-way ANOVAs indicated that individuals in the immediate condition had significantly lower Yield 1 scores than those in the delayed condition, $F(1, 158) = 8.24, p < .05, \eta^2 = .050$. Individuals in the immediate condition had significantly lower Yield 2 scores than individuals in the delayed condition, $F(1, 158) = 10.22, p < .05, \eta^2 = .061$. Individuals in the immediate condition had lower Shift scores than individuals in the delayed condition, $F(1, 158)$

= 10.98, $p < .01$, $\eta^2 = .065$. Individuals in the immediate condition had lower levels of total suggestibility than individuals in the delayed condition, $F(1, 158) = 13.47$, $p < .05$, $\eta^2 = .079$ (see Table 2).

Research question #3. How does lineup performance relate to suggestibility?

Hypothesis 8. Highly suggestible individuals will be more likely to falsely identify a witness from a target-absent lineup, as evidenced by the number of false alarms. This hypothesis was supported. A one-way ANOVA with type of interview (immediate vs. delayed, defined using a median split) as the independent variable and number of false alarms as the dependent variable revealed significant differences between these two groups, $F(1, 154) = 22.38$, $p < .01$, $\eta^2 = .127$. The average number of false alarms committed by participants who are more suggestible was 2.27 ($SD = 2.34$). The average number of false alarms committed by participants who are not as suggestible was 1.30 ($SD = 1.51$). There was a possibility of making a total of 8 false alarms. This was also shown using correlations. High scores on Total Suggestibility were related to a higher number of false alarms, $r = .397$, $p < .01$.

Hypothesis 9. Individuals who score high on Yield 2 will be more likely commit a false alarm in a target-absent lineup. This hypothesis as supported. A one-way ANOVA with Yield 2 (high vs. low, defined using a mean split) as the independent variable and number of false alarms as the dependent variable revealed significant differences between these two groups, $F(1, 154) = 20.67$, $p < .01$, $\eta^2 = .119$. The average number of false alarms committed by participants in who scored high on Yield 2 was 2.76 ($SD = 2.34$). The average number of false alarms committed by participants who scored low on Yield 2 was 1.34 ($SD = 1.56$). There was a possibility of making a total of 8 false alarms. This was also shown using correlations. High scores on Yield 2 were related to a higher number of false alarms, $r = .325$, $p < .01$.

Hypothesis 10. There will be no difference in performance between more suggestible individuals and non-suggestible individuals on number of hits. This hypothesis was supported. A one-way ANOVA with suggestibility (high vs. low) as the independent variable and number of hits as the dependent variable revealed that there were no group differences in number of hits, $F(1, 159) = .221, p > .05, \eta^2 = .639$. Total number of hits was not significantly correlated with Total Suggestibility.

To examine whether or not participants were guessing at chance level, the number of hits was divided by the total number of guesses. Participants who might have guessed at random would have correctly identified the culprits 16% of the time. Overall in the sample, participants (when they actually chose a culprit from the lineup) correctly identified the culprit 41% of the time. This did not differ according to interview type. Participants in the immediate condition did not significantly differ in the frequency of hits than participants in the delayed condition.

Additional Analyses

Interview type and suggestibility. Two-way ANOVAs were conducted to analyze the interaction between interview type and suggestibility on the free recall variables. A 2 (immediate vs. delayed) x 2 (high suggestibility vs. low suggestibility) ANOVA was conducted with total number of statements as the dependent variable. This analysis revealed that type of interview did interact with suggestibility to influence total number of statements made during free recall, $F(1, 152) = 5.45, p < .05, \eta^2 = .036$ (see Figure 1). Another 2 (immediate vs. delayed) x 2 (high suggestibility vs. low suggestibility) ANOVA was conducted with total number of accurate statements as the dependent variable. This analysis revealed that type of interview did interact with suggestibility to influence total number of accurate statements made during free recall, $F(1, 152) = 5.46, p < .05, \eta^2 = .021$ (see Figure 2). These relationships did

not hold when the number of accurate or inaccurate statements was converted to proportions of accurate or inaccurate statements. The interaction between type of interview and suggestibility did not influence total number of inaccurate statements generated, $F(1, 152) = .202, p > .05, \eta^2 = .001$.

Accuracy from Yield 1 to Yield 2. Standard error of measurement (SEM) was used to examine whether individuals increased or decreased in accuracy from Yield 1 to Yield 2. SEM allows for analysis at the individual level, rather than the group level and allows for differentiation between systematic change and unsystematic variance (Dudek, 1979; Leemrijse, Meijer, Vermeer, Lambregts, & Adèr, 1999). Using this analysis, participants' accuracy on the Yield 1 scale can be related to their accuracy score on the Yield 2 scale, as SEM can be used to estimate the unsystematic variance in an individual's score when a test is given on at least 2 occasions (Leemrijse et al.). Individuals can then be classified into three groups: a) improved in accuracy from Yield 1 to Yield 2, b) remained stable in accuracy from Yield 1 to Yield 2, or c) declined in accuracy from Yield 1 to Yield 2. The standard error of measurement analysis revealed that 42% of the sample improved in accuracy from Yield 1 to Yield 2, 30% remained stable, and 28% of the sample declined in accuracy.

Participants' classification as improved, remained stable or declined was examined according to the type of interview that they received. A chi-square analysis revealed that there were significantly more participants in the delayed group who declined in accuracy and fewer participants who increased in accuracy in comparison to the immediate group, $X^2(2, N = 160) = 11.50, p < .05$ (see Table 6).

Discussion

The current study was an investigation of eyewitness accuracy. The main goal of the study was to develop a scale to measure suggestibility in a young adult population (The Video Suggestibility Scale for Adults). This scale was developed to address the limitations of past measures of suggestibility. The second goal of the study was to examine how suggestibility might be related to misidentification. The individuals' scores on the VSSA were examined in relation to their ability to accurately select a culprit from a lineup.

The scale does appear to be working as hypothesized. Participants did yield to suggestion, but not at an extremely high rate. The mean number of yields seemed rather low in regard to total number of questions, but there was quite a lot of variability in responses. Some of the participants yielded to nearly every misleading question, while other participants only yielded to two of the misleading questions. These rates are similar to those found in the Gudjonsson Suggestibility Scales (M yield = 8.0) (Gudjonsson, 1984; 1987; Merckelbach, Muris, Wessel, & van Koppen, 1988).

The same pattern occurred for Shift. The mean number of shifts was rather low, but again, there was a lot of variability. Some participants shifted their answers to half of all of the questions, while some participants did not shift at all. Again, these rates are similar to those found in the Gudjonsson Suggestibility Scales (M Shift = 8.2) (Gudjonsson, 1984; 1987; Merckelbach, et al., 1988). Participants might not have shown high levels of shifting because the interview was so long.

Participants changed their answers significantly more when the interviewer was male than when the interviewer was female. No differences were found between interviewers for the participant's tendency to initially accept misinformation (Yield 1), but they were more likely to

change their answers (Shift) in response to the feedback provided by the male interviewer. Changing their answers resulted in a high score on the Yield 2 subscale, which Gudjonsson (1984) says is also a very good indicator of suggestibility. Differences were also found between the interviewers for Total Suggestibility. Male interviewers elicited a higher level of Total Suggestibility from participants than did female interviewers.

Gender differences in interviewer have not been found in past research with adults, so it adds to the literature. In their review of the children's suggestibility literature, Bruck, Ceci, and Melnyk (1997) discussed that when a child perceives the interviewer as being authoritative, he or she will yield to suggestions more than when the interviewer is not perceived as being an authority figure. Perhaps participants found the male interviewers to be more authoritarian and were willing to accept misinformation from him without question. Then, when he gave the negative feedback and told the participant that he or she missed many of the questions, they were intimidated into changing their answers.

Although the interaction between gender of interviewer and gender of participant was not significant, this could have occurred because the sample was overwhelmingly female. In the future, more data could be collected from males to test the interaction between gender of interviewer and gender of participant.

As previously mentioned, the means rates for Yield 1 and Shift were rather low. This could have occurred because a majority of the interviews were conducted by females (101 interviews were conducted by females, 59 interviews were conducted by males). The data show that individuals were more suggestible in response to the male interviewer than the female interviewer. Perhaps if the majority of the interviews had been conducted by a male, Yield 1 and

Shift would have been higher than if the majority of the interviews would have been conducted by females.

The mean scores for Yield 1 and Shift might not seem very high, but when placed in a real world context, these scores have far-reaching implications. Participants might have only yielded and shifted an average of 6 and 7 times (respectively), but in a real world situation this could be a tremendous barrier to the truth. If an individual agrees that he or she witnessed six different details that he or she really did not witness or if the individual changes 7 of his or her answers in response to pressure from an interviewer, this is not a reliable or accurate witness! Therefore, the percentage of questions yielded to (or the percentage of answers shifted) is not as important as the total number of yields and shifts.

The mean score for Total Suggestibility was also rather low, but again there was a lot of variability. Total suggestibility is comprised of both Yield 1 and Shift; because the mean scores of Yield 1 and Shift were rather low, Total Suggestibility was also rather low. Again this could have occurred because of the high frequency of interviews conducted by females. Female interviewers did not elicit as many Yields and Shifts as the male interviewers did. Consequently, the mean score for Total Suggestibility was not very high.

Factor Analysis of the Scale

The factor loadings do not seem to indicate that the scale is comprised of two distinct factors, Yield and Shift, as was found in the Gudjonsson scales. Keep in mind that Gudjonsson (1984) conducted his interviews shortly after the participants listened to a story about a theft and that the interview consisted of only twenty questions. This might have allowed the participants to remember what they said for each question the first round so that they could retrieve this information and change their answer when asked to do so. Therefore, there would be a more

clear distinction between the Yield and Shift subscales in Gudjonsson's scales. An item analysis could be performed that would further narrow the number of questions asked in the interview.

The VSSA interview contains 92 questions and a condition manipulation (immediate vs. delayed). Differences in participants' memories for the event might have influenced their initial Yield 1 scores. The participants had no reason to believe that they would have to remember what they said the first time, so they were probably not paying close attention to what they said. Consequently, when the interviewer provided the feedback, they had no idea what they said the first time. Therefore, there was no distinct shift from one set of answers to the next.

Narrowing down the number of questions in the interview probably would help to overcome this problem. This will have to be done before the scale can be used again. Removal of problematic items would also increase the reliabilities that were found for the Yield 1 and Yield 2 scales.

Correlates of Free Recall

Suggestibility was not associated with total number of statements generated, nor was it related to total number of accurate statements recalled. This is in line with past research (Ebbesen & Rienick, 1982; Eugenio et al., 1982; Yuille & Cutshall, 1986) which found that individuals are generally accurate when asked to recall details. Number of accurate statements recalled was negatively related to scores on the Yield 1 subscale. Participants who had a high number of accurate statements in their free recall were less likely to initially accept misinformation from the misleading questions. This varied by interview type, with participants in the immediate condition performing much better than the other groups. Participants in the immediate condition who were less suggestible had the highest numbers of accurate statements

in their free recall in comparison to participants who were more suggestible or those who were in the delayed condition.

Yield 2 scores and Total Suggestibility scores were related to higher numbers of total number of inaccurate statements recalled. This must be interpreted with some caution, as only a small percentage of total number of statements recalled were inaccurate statements. In general, individuals were quite accurate in their recall of the events.

Interesting patterns emerged among the subscales of the VSSA and the free recall components. Participants who scored high on the Yield 1 scale also had a lower number of accurate statements than participants who scored low on Yield 1. Participants who scored high on Yield 1 might not have generated a high number of accurate statements because they did not pay close attention to the video. If this is the case, they would not have recalled a high number of accurate details and they would have scored high on Yield 1 because they simply did not know the correct answer. They might have accepted the misinformation because they were guessing. This is in line with an explanation that Loftus and Hoffman (1989) give about how misinformation is introduced into memory; the stimulus is never encoded, so the participant guesses an answer instead of simply saying “I don’t know.” Furthermore, if he or she was not paying attention, he or she might have agreed with what the interviewer was asking in an attempt to “prove” that they were paying attention.

Immediate versus Delayed Interviews

Free recall. Timing of interview did appear to affect free recall. Participants in the immediate condition recalled a higher number of total statements and a higher number of accurate statements than the participants in the delayed condition. There were no differences found for total number of inaccurate statements. The explanation for the first two findings is

clear; memory decays over time (Kassin et al., 1989), so number of total statements recalled would probably decrease as time goes by. The participants in the immediate condition had seen the video only about 30 minutes earlier, so the video was still fresh in their minds. Because they had just recently seen the video, they were able to generate quite a large number of accurate statements. Participants in the delayed condition had seen the video 5 to 8 days before the interview. Their memories for the event declined over time and they were not able to recall a large number of details.

The groups might not have differed on total number of inaccurate statements because the video was too short and the events were easy to follow. With an average of only two inaccurate statements given in free recall, participants were performing at floor level for number of inaccurate statements. Perhaps a longer video would lead to larger differences between the immediate and delayed conditions. The free recall portion of the video was administered before the introduction of the misinformation. If the free recall was assessed after the misinformation, there might have been larger differences between the immediate and delayed conditions as well.

Suggestibility. Timing of interview did have an effect on the VSSA subscales. Participants in the immediate condition had lower Yield 1, Yield 2, Shift, and Total Suggestibility scores in comparison to participants in the delayed condition. In the immediate condition, participants had just seen the video and were not easily misled by the misinformation supplied by the interviewer. Furthermore, these participants might have been fairly confident in their answers and were not fooled when the interviewer said that they missed a lot of the questions. Therefore, they were not likely to change their answers very often, resulting in lower Yield 2 scores; if they felt they were correct the first time the question was asked, they would not have changed their answers very much. However, individuals in the delayed condition might

have been not as confident in their initial answers and were likely to believe the interviewer when told they had missed a lot of the questions. Consequently, they would change their answers more than the participants in the immediate condition, resulting in higher Shift scores.

The differences between the groups for Total Suggestibility can most likely be explained by the differences in Yield 1 and Shift. Because Total Suggestibility is comprised of the sum of Yield and Shift, the delayed group with their higher Yield 1 and Shift scores resulted in higher Total Suggestibility scores. It is interesting to note that the differences in Yield 1 scores between participants in each condition was small, but significant (6.0 for the immediate condition vs. 7.6 for the delayed condition). The difference between the groups on Shift, however, is larger than the difference found between the groups for shift. This is an important finding as it shows that individuals might have a fairly good memory for an event, even when asked misleading questions, but that after a delay, individuals are more easily swayed when pressured to change their answers than when they are interviewed immediately.

Lineup Performance

The current study found a relationship between suggestibility and one's ability to correctly identify a culprit from a lineup. Individuals who were more suggestible had a higher mean number of false alarms than individuals who were not as suggestible. Based on the methods used in this study, it can be concluded that participants who are more suggestible do not form solid memories for faces; rather, participants who are more suggestible might be easily convinced into believing that they saw a face when in reality they did not. Further analysis of the lineups can describe whether or not the participants are committing false alarms because they are selecting the foils that have been repeated throughout the lineups or whether they are simply guessing incorrectly.

Furthermore, because compliance is related to Yield 1 (a component of suggestibility) these participants might have been more likely to comply with the interviewer and select an individual from a lineup, regardless of whether or not they believed the perpetrator was present.

There was no relationship between suggestibility and total number of hits. It appears that when an individual is faced with the correct stimulus (i.e., the actual perpetrator), suggestibility has little to do with their ability to correctly recognize that stimulus. However, the average number of hits was only 1.2 out of a possible 4 hits. The conclusion is that suggestibility did not have an influence over participants' ability to correctly choose a perpetrator from a target-present lineup because all participants performed near floor level.

Timing of interview and suggestibility. Interesting results were found for the interaction between timing of interview and one's level of suggestibility. For both total number of statements generated and total number of accurate statements, participants in the immediate condition who were less suggestible generated the highest, followed by participants in the immediate condition who are suggestible, participants in the delayed condition who were more suggestible, and participants in the delayed condition who were less suggestible.

There were no significant differences between participants who were more suggestible and participants who were less suggestible on total number of statements or total number of accurate statements generated. However, timing of interview did interact with suggestibility to influence these variables. There was a main effect present for type of interview, with participants in the immediate group generating more statements and more accurate statements than those in the delayed group. Furthermore, among those in the immediate group, participants who were less suggestible generated significantly more statements and more accurate statements than those who were more suggestible.

Accuracy from Yield 1 to Yield 2. In the total sample, there were many people who improved in accuracy from Yield 1 to Yield 2. Improvement was defined as changing incorrect responses during Yield 1 questioning to correct responses during the Yield 2 questioning. Therefore, the Yield 2 score was at least one standard error of measurement higher than the Yield 1 score. Decline was defined as changing many correct responses during Yield 1 questioning to incorrect answers during Yield 2 questioning. In this category, the Yield 2 score was at least one standard error of measurement lower than the Yield 1 score. Those who were classified as stable did not significantly change from Yield 1 to Yield 2. These Yield 2 scores for these participants were within one standard error of measurement above or below their Yield 1 scores.

When the classifications were examined in terms of type of interview, very different patterns emerged for the two groups. Participants in the immediate condition most often improved from Yield 1 to Yield 2, and only a few of the participants declined in accuracy from Yield 1 to Yield 2. The opposite pattern emerged in the delayed group. These participants most often declined in accuracy and less than half as many participants in the delayed condition improved in comparison to those in the immediate condition.

This is complementary to the findings that only small (but significant) differences were found between the two types of interviews on Yield 1, but large differences were found for Shift. Both groups had fairly low Yield 1 scores, but the discrepancies between Shift were very large. If the groups started out the nearly the same, but participants in the delayed condition changed their answers more often than participants in the immediate condition they would therefore have higher Shift scores. Participants in the delayed condition clearly changed their answers at a higher rate than participants in the immediate condition. Because they started out similar to

participants in the immediate condition, it follows that because of the high frequency of changed answers, more of the participants in the delayed condition declined in accuracy from Yield 1 to Yield 2 in comparison to the immediate group.

Limitations and Future Directions

The current study utilized undergraduates from West Virginia University. Although the sample might have been representative of the population of West Virginia, the sample was not representative of the entire population. The majority of the sample was White women in their late teenage years or early twenties. This was a limitation as firm conclusions cannot be made about race or gender of the participant. Future research should include more even numbers of males and females in the study to draw firm conclusions about gender differences in suggestibility. Also, age differences were not well represented. It is unclear whether or not the scale will work with younger, adolescent populations or middle-to-older adult populations. Future research could address this by using different aged samples.

Furthermore, the sample consisted of all volunteers who sought extra credit for participation. They might not have taken the study seriously and were only concerned about finishing the procedure and receiving their extra credit. On a few occasions, participants were observed trying to check their email on the computer or looking through their backpacks while the video was on. Although these limitations were overcome after the discovery of these behaviors, this is an obvious limitation to the study.

Another limitation is that there were several different interviewers. Each interviewer might have (despite training) developed their own interviewing style. Participant might have reacted to the interviewer's unique personality, causing differences in performance. The sample is too small to analyze the influence of each interviewer; however, gender of interviewer did

make a difference. Future research should use one male interviewer and one female interviewer to further explore the influence of gender of interviewer. Interactions between gender of interviewer and gender of participant could also be examined with equal numbers.

The video has a very amateur-like quality about it, and is often very shaky and hard to follow. Some participants commented on how it was rather unsettling to watch and one participant actually asked for a Pepto Bismol after viewing the video. Due to the shaky quality of the video, many participants might not have been able to follow each step of the video or get a good look at the perpetrators. However, the video was of good enough quality for most of the participants to get the gist of the video, as evidenced by their performance in the free recall portion.

Although many studies look at memory for a stimulus such as a video or a story, the situation is somewhat artificial. This issue was circumvented somewhat by telling the participants that these videos were videos of actual crimes that were committed by young men on a crime spree. The actors in the video were not professionals and at times it is apparent that the events were staged. Because of this, participants might not have taken the video seriously. Future research could examine the validity of the procedures by examining how well the video relates to suggestibility about a live, witnessed event.

To further examine the relationship between lineup performance and suggestibility, future research should address overall memory ability and compliance. These two variables might moderate the relationship between suggestibility and misidentification. Because Yield is related to compliance (Richardson & Kelly, 2004), compliance might also influence an individual's tendency to select a perpetrator's face from the lineup. A compliant individual might want to please the interviewer and will select a face simply because they feel that they

have to, regardless of his or her suggestibility. Noncompliant individuals might not select anyone from a lineup if they do not feel confident that the person is there, regardless of his or her level of suggestibility.

Conclusions and Real-World Applications

The current study presented a new measurement to capture suggestibility in a young adult population. This measure shows that individuals are generally accurate when they are asked to recall events that they witnessed, but that when asked misleading probe questions, they often give in to misinformation. In particular, the current study showed that even a one-week delay can have detrimental effects on individuals' memories. After a one-week delay, participants were less likely to generate accurate statements in the free recall and they were much more suggestible than participants who were interviewed immediately. Furthermore, the current study showed that individuals who are more suggestible are less likely to correctly identify a culprit from a target-present lineup than individuals who are less suggestible.

Law enforcement personnel should be aware that the time elapsed since witnessing an event is an important factor to consider. They should be aware that the sooner the witness is interviewed, the more accurate their statement will probably be. Also, if a long period of time has elapsed between witnessing the event and being interviewed, individuals are more likely to accept false suggestion. This means that law enforcement personnel should be very careful about how they word their questions and be careful to not lead witnesses, particularly if the witness is being interviewed after a long delay.

The Video Suggestibility Scale for Adults could be used in a real world setting. Although it is not intended to be a tool to determine who should be allowed to testify or not, it might be useful in its ability to determine who is most likely to accept false suggestion.

Table 1

Demographic Characteristics of the Interview Groups

	Immediate	Delayed	Never Returned
Age (mean)	19.9 (1.5)	19.3 (1.4)	20.7 (3.7)
GPA (mean)	3.0 (0.7)	3.2 (0.6)	3.1 (0.5)
Gender (number)			
Male	18	9	4
Female	68	65	17
Race (number)			
Caucasian	69	69	20
African American	3	3	1
Native American	1	0	0
Asian American	9	0	0
Other	4	2	0
Class Rank (number)			
Freshman	25	30	3
Sophomore	19	25	12
Junior	24	11	3
Senior	17	8	3

Note: standard deviation in parentheses.

Table 2

Means on all of the Measures by Interview Group

	Immediate	Delayed
Yield 1	6.0 (3.1)	7.6 (4.1)
Yield 2	7.5 (5.0)	10.3 (5.9)
Shift	10.3 (7.3)	14.2 (8.7)
Total Suggestibility	17.1 (9.1)	23.1 (11.3)
Total False Alarms	1.7 (1.9)	2.1 (2.1)
Total Hits	1.3 (0.8)	1.1 (1.0)

Note: standard deviation in parentheses.

Table 3

Correlations Among the VSSA Subscales and the Free Recall Components

	Yield 1	Yield 2	Shift	TSugg	TState	Acc	Inacc
Yield 1	1						
Yield 2	.660 **	1					
Shift	.488 **	.807 **	1				
TSugg	.733 **	.863 **	.951 **	1			
TState	-.154	-.079	-.068	-.106	1		
Acc	-.183 *	-.110	-.106	-.145	.989 **	1	
Inacc	.127	.161 *	.217 **	.213 **	.322 **	.179 *	1

Note: TSugg = Total Suggestibility, TState = Total number of statements generated, Acc. = total number of accurate statements generated, Inacc. = Total number of inaccurate statements generated.

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

Table 4

Sex of Interviewer and the VSSA Subscales

Sex of Interviewer	Yield 1	Yield 2**	Shift**	Total Suggestibility**
Male	7.2 (4.0)	10.1 (6.6)	15.5 (9.4)	23.7 (12.3)
Female	6.4 (3.5)	8.0 (4.8)	10.3 (6.8)	17.7 (8.8)

Note: ** $p < .01$

Table 5

Means of Free Recall Variables by Condition

Condition	Total Statements*	Total Accurate Statements*	Total Inaccurate Statements
Immediate	35.4 (17.6)	32.7 (17.1)	2.6 (2.0)
Delayed	25.4 (10.1)	22.8 (9.0)	2.5 (2.6)

Note: standard deviation in parentheses.

Note: * $p < .05$

Table 6

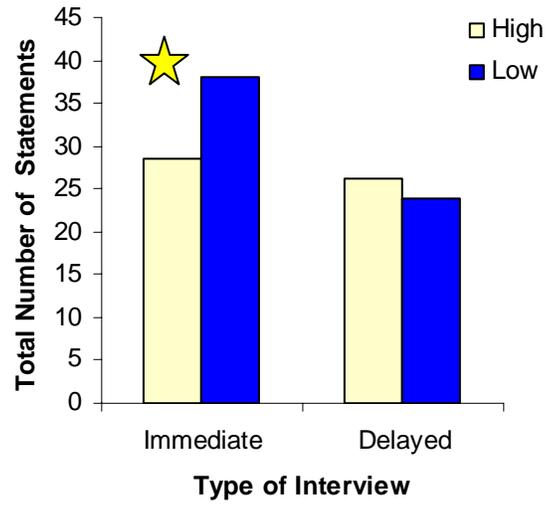
Frequencies of Participants who Improved, Remained Stable, or Declined in Accuracy from Yield 1 to Yield 2

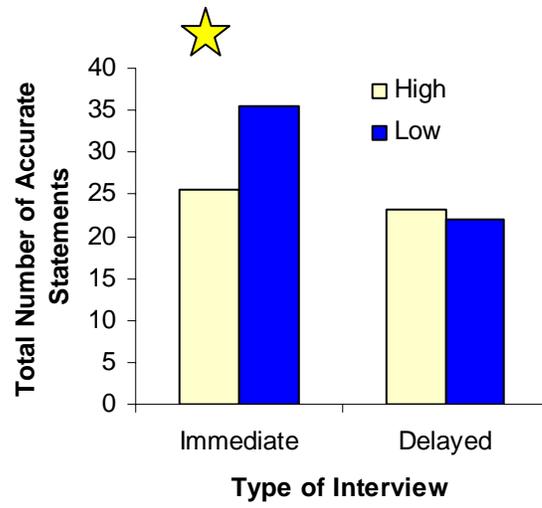
Type of Interview	Improved	Remained Stable	Declined
Immediate	47	22	17
Delayed	21	27	26

Figure Captions

Figure 1. The interaction between type of interview and suggestibility on total number of statements generated in free recall.

Figure 2. The interaction between type of interview and suggestibility on total number of accurate statements generated in free recall.





References

- Baxter, J. S., & Boon, J. C. W. (2000). Interrogative suggestibility: The importance of being earnest. *Personality and Individual Differences, 28*, 753-762.
- Bruck, M., Ceci, S., & Melnyk, L. (1997). External and internal sources of variation in the creation of false reports in children. *Personality and Individual Differences, 9*, 289-316.
- Clare, I. C. H., & Gudjonsson, G. H. (1993). Interrogative suggestibility, confabulation, and acquiescence in people with mild learning disabilities (mental handicap): Implications for reliability during police interrogations. *British Journal of Clinical Psychology, 32*, 295-301.
- Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology, 66*, 364-370.
- Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology, 63*, 352-359.
- Dudek, F. J. (1979). The continuing misinterpretation of the standard error of measurement. *Psychological Bulletin, 86*, 335 – 337.
- Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83*, 745-762.
- Eugenio, P., Buckhout, R., & Kostas, S. (1982). Hypermnnesia in the eyewitness to a crime. *Bulletin of the Psychonomic Society, 19*, 83-86.
- Gudjonsson, G. H. (1983). Suggestibility, intelligence, memory recall and personality: An experimental study. *British Journal of Psychiatry, 142*, 35-37.
- Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences, 5*, 303-314.

- Gudjonsson, G. H. (1986). The relationship between interrogative suggestibility and acquiescence: Empirical findings and theoretical implications. *Personality and Individual Differences, 7*, 195-199.
- Gudjonsson, G. H. (1987). A parallel form of the Gudjonsson suggestibility scale. *British Journal of Clinical Psychology, 26*, 215-221.
- Gudjonsson, G. H. (1987a). Historical background to suggestibility: How interrogative suggestibility differs from other types of suggestibility. *Personality and Individual Differences, 8*, 347-355.
- Gudjonsson, G. H. (1988). Interrogative suggestibility: Its relationship with assertiveness, social-evaluative anxiety, state anxiety and method of coping. *British Journal of Clinical Psychology, 27*, 159-166.
- Gudjonsson, G. H. (1991). The effects of intelligence and memory on group differences in suggestibility and compliance. *Personality and Individual Differences, 12*, 503-505.
- Gudjonsson, G. H. (1991). The effect of intelligence and memory on group differences in suggestibility and compliance. *Personality and Individual Differences, 12*, 503-505.
- Gudjonsson, G. H. (1992) Interrogative suggestibility: Factor analysis of the Gudjonsson suggestibility scale. *Personality and Individual Differences, 13*, 479-481.
- Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. Hoboken, NJ: Wiley.
- Gudjonsson, G. H., & Clare, I. C. H. (1995). The relationship between confabulation and intellectual ability, memory, interrogative suggestibility, and acquiescence. *Personality and Individual Differences, 10*, 333-338.

- Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour, 1*, 83-104.
- Gudjonsson, G. H., Rutter, S. C., & Clare, I. C. H. (1995). The relationship between suggestibility and anxiety among suspects detained at police stations. *Psychological Medicine, 25*, 875-878.
- Gudjonsson, G. H., & Sigurdsson, J. F. (1996). The relationship of confabulation to the memory, intelligence, suggestibility and personality of prison inmates. *Applied Cognitive Psychology, 10*, 85-92.
- Gudjonsson, G. H., & Singh, K. K. (1984). Interrogative suggestibility and delinquent boys: An empirical validation study. *Personality and Individual Differences, 5*, 425-430.
- Kassin, S. M. (1998). Eyewitness identification procedures: The fifth rule. *Law and Human Behavior, 22*, 649-653.
- Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The “general acceptance” of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 44*, 1089-1098.
- Leemrijse, C., Meijer, O. G., Vermeer, A., Lambregts, B., & Adèr, H.J. (1999). Detecting individual change in children with mild to moderate motor impairment: the standard error of measurement of the Movement ABC. *Clinical Rehabilitation, 13*, 420 – 430.
- Liebman, J. I., McKinley-Pace, M. J., Leonard, A. M., Sheesley, L. A., Gallant, C. L., Renkey, M. E., & Lehman, E. B. (2002). Cognitive and psychosocial correlates of adults’ eyewitness accuracy and suggestibility. *Personality and Individual Differences, 33*, 49-66.

- Loftus, E. F., & Burns, T. E. (1982). Mental shock can produce retrograde amnesia. *Memory and Cognition*, *10*, 318-323.
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*, 560-572.
- Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General*, *118*, 100-104.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 19-31.
- Loftus, E. F., & Palmer, J. C. (1984). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 585-589.
- Merckelbach, H., Muris, P., Wessel, I., & van Koppen, P. J. (1998). The Gudjonsson Suggestibility Scale (GSS): Further information on its reliability, validity, and metacognition correlates. *Social Behavior and Personality*, *26*, 203-210.
- Miguel, M., Garcia-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology*, *13*, 257-268.
- Prkachin, G. C. (2003). The effects of orientation on detection and identification of facial expressions of emotion. *British Journal of Psychology*, *94*, 45-62.
- Richardson, G., Gudjonsson, G. H., & Kelly, T. P. (1995). Interrogative suggestibility in an adolescent forensic population. *Journal of Adolescence*, *18*, 211-216.

- Richardson, G., & Kelly, T. P. (2004). A study in the relationship between interrogative suggestibility, compliance, and social desirability in institutionalized children. *Personality & Individual Differences, 36*, 485-494.
- Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnesia for details about a violent event. *Journal of Applied Psychology, 73*, 371-377.
- Sigurdsson, J. F., & Gudjonsson, G. H. (1996). The psychological characteristics of 'false confessors'. A study among Icelandic prison inmates and juvenile offenders. *Personality and Individual Differences, 20*, 321-329.
- Tollestrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144-160). New York: Cambridge University Press.
- Turtle, J. W., & Yuille, J. C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology, 79*, 260-271.
- Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36*, 1546-1557.
- Wells, G. L. (1993). What do we know about eyewitness testimony? *American Psychologist, 48*, 553-571.
- Wells, G.L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295.
- Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22*, 603-?.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71*, 291-301.

Appendix A

Demographic Questionnaire

1. What is your sex?
 male
 female
2. What is your age?
 years
3. What is your class rank?
 freshman
 sophomore
 junior
 senior
 graduate
4. What is your overall GPA?
5. What is your ethnic background?
 Caucasian
 African American
 Asian American
 Native American
 Middle Eastern
 Other
6. How would you rate your overall memory ability?
 much better than average
 somewhat better than average
 average
 somewhat below average
 much below average
7. Are you familiar with the procedures of a police lineup?
 yes
 no

Appendix B

The Video Suggestibility Scale for Adults

Eyewitness Memory

One to three days after the individual has seen the video:

Open-ended questions

- A. Do you remember the video about the carjacking? Tell me everything you remember about the carjacking, even the things you think aren't important.
*Repeatedly probe: **What else do you remember?***

When the individual finishes responding, ask?

B. Who was at the carjacking? *For each character from the list of characters that the individual generates AND for any characters mentioned above in the open-ended questioning, ask the following two questions:*

B1a. What did _____ look like?

B1b. What was he/she wearing?

B2a. What did _____ look like?

B2b. What was he/she wearing?

B3a. What did _____ look like?

B3b. What was he/she wearing?

C. Continue asking about characters generated by the individual and conclude by asking: **Who else was at the carjacking?**

C1a. What did _____ look like?

C1b. What was he/she wearing?

- C2a. What did _____ look like?
C2b. What was he/she wearing?

Open-ended questions

- D. Do you remember the video about the burglary? Tell me everything you remember about the burglary, even the things you think aren't important.
*Repeatedly probe: **What else do you remember?***

When the individual finishes responding, ask?

E. Who was at the burglary? *For each character from the list of characters that the individual generates AND for any characters mentioned above in the open-ended questioning, ask the following two questions:*

- E1a. What did _____ look like?
E1b. What was he/she wearing?

- E2a. What did _____ look like?
E2b. What was he/she wearing?

- E3a. What did _____ look like?
E3b. What was he/she wearing?

F. Continue asking about characters generated by the individual and conclude by asking:
Who else was at the burglary?

- F1a. What did _____ look like?
F1b. What was he/she wearing?

- F2a. What did _____ look like?
F2b. What was he/she wearing?

G. Probe questions *I am going to ask you some questions about the carjacking.*

1. Was the perpetrator wearing a baseball hat?
2. Was the perpetrator wearing a dark blue sweatshirt?
3. Was the perpetrator bald?
4. Was the victim wearing shorts?
5. Was the victim blonde?
6. Was the victim wearing a green shirt?
7. Was the victim wearing glasses?
8. Did the perpetrator have a knife?
9. Were there any witnesses?
10. Was there a highway nearby?
11. Were there houses nearby?
12. Was the victim's car the only one in the parking lot?
13. Was the camera that the perpetrators used stolen?
14. Were the perpetrators in a parking lot?
15. Was it the cameraman's idea to steal a car?
16. Were the perpetrators admiring the car before they saw the woman?
17. Were the perpetrators waiting for the girl to leave the store?
18. Was the woman leaving a CVS store?
19. Did the perpetrator leave behind the knife?
20. Was the victim talking on her cell phone?
21. Did the victim speak to the perpetrators?
22. Did the victim drop her purse?
23. Did the victim scream for help?
24. Did the victim attempt to use her mace?
25. Did the victim's friend run away?
26. Did the victim run toward her vehicle?
27. Did you hear the baseball bat hit the victim's head?
28. Did the bat get bloody?
29. Did the perpetrator kick the victim when she was on the ground?
30. Did the perpetrator drop the bat he used on the victim?
31. Did the victim's shoes fall off when she fell to the ground?
32. Did the victim tear her dress?
33. Did the perpetrator stab the victim?
34. Did the perpetrators hit the girl with the car?
35. Did the perpetrator hotwire the car?
36. Did the perpetrators steal the victim's purse?
37. Did the perpetrators steal the victim's wallet?
38. Was the victim coherent and moving when the perpetrators fled the scene?
39. Did the perpetrator take her necklace?
40. Did the perpetrators run into an alley?
41. Were you able to hear sirens in the background?
42. Did the perpetrators leave the purse behind after they abandoned the car?
43. Were the perpetrators laughing about the crime?

(go on the ask questions about the burglary)

H. Probe questions *Now I am going to ask you some questions about the burglary.*

1. Did one of the burglars have curly hair?
2. Did one of the burglars have a goatee?
3. Was one of the burglars clean shaven?
4. Did the cameraman have a black shirt on?
5. Was the girl wearing a skirt?
6. Did the little girl have black hair?
7. Did the man with the curly hair have glasses?
8. Did one of the burglars smoke?
9. Was one of the burglars wearing a hat and glasses?
10. Did the cameraman break into the house once before?
11. Did the burglars enter into the kitchen of the house?
12. Did the burglars take two beers out of the refrigerator?
13. Did the burglars turn off the burglar alarm?
14. Was the computer turned on?
15. Were the burglars drinking beer while robbing the house?
16. Did one of the burglars have a fear of dogs?
17. Were the homeowners not home because they were on vacation?
18. Were the neighbors at the movies that night?
19. Did they steal video games?
20. Did they steal a bass guitar?
21. Did they steal some jewelry?
22. Did the burglars hear police sirens?
23. Did the burglars exit through a window?
24. Did you see the burglar place the *Palm Pilot* into his pocket?
25. Were you able to hear traffic in the background?
26. Were the burglars concerned about the traffic?
27. Did one of the burglars trip and fall when he was going upstairs?
28. Were the lights on in the living room when they first entered the house?
29. Did you see the burglar put the jewelry box into his pocket?
30. Were the burglars startled when they heard the knock on the door?
31. Did one of the burglars previously cut the screen in order to gain entry?
32. Were the burglars armed with guns?
33. Was the dog outside on its runner?
34. Did the owners of the house yell at the burglars as they were leaving?
35. Did the burglars steal the car keys?
36. Did one of the burglars check his email?
37. Did they steal the computer?
38. Did the burglars steal money out of the wallet?
39. Did they steal the TV so they could play the video games?
40. Did they take the *Play Station*?
41. Was one of the burglars bleeding after he fell down the stairs on the way out?
42. Did the little girl wear glasses?
43. Did the little girl call the police?
44. Did the little girl yell for her mom when she caught the perpetrators?
45. Did all of the burglars escape the scene safely?

46. Did the burglars leave anything behind at the scene?
47. Were the burglars playing the video games they stole?
48. Did they decide to rob another house?
49. Did the cameraman say that he would make copies of the tape for the other burglars?

(Give feedback)

Appendix C

Lineup Procedures

Now I am going to show you some lineups.

Lineup A

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup B

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup C

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup D

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup E

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup F

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup G

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup H

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup I

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup J

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup K

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9

Lineup L

For this lineup, I want you to examine each face and tell me if you recognize anyone from the video.

Individual selects:

No one _____

Person in position _____

Confidence 1 2 3 4 5 6 7 8 9