Florida International University
FIU Digital Commons

FIU Electronic Theses and Dissertations | University Graduate School

8-11-2008

Eyewitness Memory: Balancing the Accuracy, Precision, and Quantity of Information through Metacognitive Monitoring and Control

Jacqueline Evans
Florida International University

DOI: 10.25148/etd.FI10022510
Follow this and additional works at: https://digitalcommons.fiu.edu/etd

Recommended Citation
Evans, Jacqueline, "Eyewitness Memory: Balancing the Accuracy, Precision, and Quantity of Information through Metacognitive Monitoring and Control" (2008). *FIU Electronic Theses and Dissertations*. 188.
https://digitalcommons.fiu.edu/etd/188

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

EYEWITNESS MEMORY: BALANCING THE ACCURACY, PRECISION, AND QUANTITY OF INFORMATION THROUGH METACOGNITIVE MONITORING AND CONTROL

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Jacqueline Evans

2008

To: Dean Kenneth Furton
College of Arts and Sciences

This dissertation, written by Jacqueline Evans, and entitled Eyewitness Memory: Balancing the Accuracy, Precision, and Quantity of Information through Metacognitive Monitoring and Control, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Janat Parker

_____
Nadja Schreiber Compo

_____
Howard Wasserman

_____
Ronald Fisher, Major Professor

Date of Defense: August 11, 2008

The dissertation of Jacqueline Evans is approved.

_____
Dean Kenneth Furton
College of Arts and Sciences

_____
Dean George Walker
University Graduate School

Florida International University, 2008

DEDICATION

I dedicate this dissertation to everyone who has helped to make its completion possible, and to everyone who was there for me on the days when it seemed its completion was in fact impossible.

ACKNOWLEDGMENTS

I would like to thank my committee members for their extreme flexibility and support throughout this process. I would also like to acknowledge the research assistants who put in many hours and made this project possible: Giancarlo Galliani, Diana Garcia, Araceli Leyva, Piero Peirano, Stephanie Roque & Dominique Stauffer. I would also like to acknowledge Dr. Nadja Schreiber Compo for allowing me to use both her lab space and research assistants in order to collect the data for this dissertation.

ABSTRACT OF THE DISSERTATION

EYEWITNESS MEMORY: BALANCING THE ACCURACY, PRECISION, AND QUANTITY OF INFORMATION THROUGH METACOGNITIVE MONITORING AND CONTROL

by

Jacqueline Evans

Florida International University, 2008

Miami, Florida

Professor Ronald Fisher, Major Professor

Historically, memory has been evaluated by examining how much is remembered, however a more recent conception of memory focuses on the accuracy of memories. When using this accuracy-oriented conception of memory, unlike with the quantity-oriented approach, memory does not always deteriorate over time. A possible explanation for this seemingly surprising finding lies in the metacognitive processes of monitoring and control. Use of these processes allows people to withhold responses of which they are unsure, or to adjust the precision of responses to a level that is broad enough to be correct. The ability to accurately report memories has implications for investigators who interview witnesses to crimes, and those who evaluate witness testimony.

This research examined the amount of information provided, accuracy, and precision of responses provided during immediate and delayed interviews about a videotaped mock crime. The interview format was manipulated such that a single free narrative response was elicited, or a series of either yes/no or cued questions were asked. Instructions provided by the interviewer indicated to the participants that they should

either stress being informative, or being accurate. The interviews were then transcribed and scored.

Results indicate that accuracy rates remained stable and high after a one week delay. Compared to those interviewed immediately, after a delay participants provided less information and responses that were less precise. Participants in the free narrative condition were the most accurate. Participants in the cued questions condition provided the most precise responses. Participants in the yes/no questions condition were most likely to say "I don't know". The results indicate that people are able to monitor their memories and modify their reports to maintain high accuracy. When control over precision was not possible, such as in the yes/no condition, people said "I don't know" to maintain accuracy. However when withholding responses and adjusting precision were both possible, people utilized both methods. It seems that concerns that memories reported after a long retention interval might be inaccurate are unfounded.

TABLE OF CONTENTS

Chapter I

Literature Review

Approaches to Memory Research: Quantity versus Accuracy

Memory researchers can go about their task in several ways. Historically, they have tended to use list learning tasks to measure memory performance. In a typical list learning task, subjects are exposed to a list of words and after a specified amount of time has passed they are asked to recall as many words as possible. Researchers using such tasks are often interested only in one aspect of the subjects' performance: *how many* words they can correctly recall (Koriat & Goldsmith, 1994). This measure of memory is input-bound; memory is assessed in terms of the percentage recalled out of the total amount presented at encoding (the input into memory). Those who use such quantity-oriented measures of memory are often conceptualizing memory as a kind of storeroom where information is kept until it needs to be retrieved. While this approach to memory research has been adopted by many researchers, it is not the only conception available.

Another approach to memory research is more concerned with the *accuracy rate* of what is reported during recall, as opposed to how much is reported. Accuracy rate is an output-bound measure of memory, as performance is assessed by examining the proportion of accurate information reported (the output from memory). This accuracy-oriented approach to memory is much more valuable in some contexts than the quantity-oriented approach. For example, when eyewitnesses to a crime are asked to recount what they saw, even more important than the absolute number of correct details they remember is to what extent the reported information can be trusted to be correct. Essentially, what witnesses forget is not of primary concern, but what they remember and report is critical.

Of course, more information is more helpful than less information, and researchers are interested in finding ways to obtain more information from witnesses. However for any increase in information quantity to be valuable a high accuracy rate must be maintained (see Fisher & Geiselman, 1992).

Quantity and Accuracy Rates across Retention Intervals

One of the classic findings in psychology is the forgetting curve (Ebbinghaus, 1895/1964); this curve begins with a sharp memory decline in the hours after a target event and then flattens as time passes. However, this forgetting curve oversimplifies the effect of time on memory retrieval. First, this approach is strictly quantitative in nature, and thus tells us nothing about accuracy rates. Also, studies examining memory over time have shown that memory for fine grained information deteriorates relatively rapidly, while memory for more coarse grained information endures (e.g., Christiaansen, 1980; Conway, Cohen, & Stanhope, 1991; Koriat, Levy-Sadot, Edry, & de Marcas, 2003; Stanhope, Cohen, & Conway, 1993). For example, memory for gist, or general meaning, endures longer than memory for surface form and verbatim information (e.g., Kintsch, Welsch, Schmalhofer, & Zimmy, 1990). There is also evidence that memory for categories is retained longer than memory for items (Dorfman & Mandler, 1994), and that there is a retention advantage of basic level information (Pansky & Koriat, 2004). As memory for different levels of information declines at different rates, it is not possible to simply apply the forgetting curve to memory for complex information (Goldsmith, Koriat, & Pansky, 2005).

Eyewitness Memory Accuracy and Metacognitive Monitoring and Control

Long-term retention of memory for complex events is very relevant to research examining eyewitness memory. While witnesses may (or may not) be interviewed by the police soon after a crime takes place, much more time elapses between that interview with the police and witnesses' testimony in court. It would seem reasonable to assume that witnesses would be less accurate after such long retention intervals than after relatively short retention intervals. Interestingly, this often does not seem to be the case; researchers have found unexpectedly stable accuracy rates for memories of events at retention intervals from one month (e.g., Ebbesen & Rienick, 1998) up to several years (e.g., Poole & White, 1993). Thus, accuracy rates and quantity are not affected in the same way by the passage of time.

There are two mechanisms that have been examined which help explain how memory accuracy can be maintained from immediate tests across prolonged retention intervals: choosing not to provide a response that is likely to be incorrect (exercising a report option), and adjusting a response's grain size to a level that is likely to be correct (control over precision). These two mechanisms will be discussed in detail below, but first the role of metacognition in maintaining accuracy is discussed.

Good metacognitive skills are necessary for individuals to effectively use their options to withhold a response and to control response precision; specifically, both effective monitoring and control processes are required (Koriat & Goldsmith, 1996). Monitoring is the ability to accurately estimate the probability that a response is correct. When monitoring is effective there is a strong positive correlation between confidence in a given response and the accuracy of that response; if monitoring is ineffective there is

either no correlation, or an inverse correlation. Control is the ability to successfully filter responses based on a pre-set, but flexible, response criterion that determines if an answer should be volunteered or not. The decision to respond is made after comparing confidence in a response with the pre-set response criterion. If confidence in a response is higher than the response criterion the response should be volunteered. If confidence in a response is lower than the response criterion then the response should be withheld.

The Role of Report Option

Koriat and Goldsmith (1996) provided a framework describing how a report option can improve accuracy through a quantity-accuracy trade-off. Simply put, when people are not confident in a response they may simply choose to refrain from answering the question. By systematically choosing not to answer questions when they are not confident in the answer, as long as monitoring is effective and control is sensitive, their overall accuracy rate will increase since only answers that are likely to be correct are volunteered. Of course choosing not to answer questions means that this higher accuracy rate was bought at the cost of information quantity. Several studies have found evidence of such a quantity-accuracy trade-off (e.g. Koriat & Goldsmith, 1994, 1996; Kelley & Sahakyen, 2003).

In the first study examining the role of report option on accuracy maintenance, Koriat and Goldsmith (1994) fully crossed report option (free v. forced response) and test format (recall v. recognition). This experiment is unique in that it unconfounds report option and test format; most studies have looked at only (free) recall and (forced) recognition, which resulted in a recall-recognition paradox. The recall-recognition paradox describes the two conflicting views supported by memory research: recall

memory is better than recognition memory because the accuracy rate of recall reports is higher than the accuracy rate of recognition tests, yet recognition memory is superior to recall memory because the quantity of correct information provided in recognition tests is larger than the quantity of correct information provided in recall tests (e.g., Lipton, 1977). Koriat and Goldsmith's study revealed the explanation for this paradox: test format has usually been confounded with type of memory task.

Koriat and Goldsmith (1994) found in their first experiment that report option was successfully utilized when it was available: participants were more accurate in the free report conditions than the forced report conditions because they were able to screen out answers that were unlikely to be correct. Importantly, accuracy rates were the same for recall and recognition in the free report condition, indicating that report option was the key variable determining accuracy rate, not test format. In their third experiment Koriat and Goldsmith demonstrated that accuracy was under strategic control, as manipulating the incentive to provide correct information resulted in accuracy rate adjustments. In the first experiment, described above, the participants were told they would earn approximately $0.25 for each correct answer, and would lose the same amount if they provided an incorrect answer. In contrast, in the third experiment the participants were told that they would earn approximately $0.25 for each correct answer, but if they provided a single incorrect answer they would lose all of their earnings. In the third study (the high incentive condition) the accuracy rate for the free report group was much higher than the accuracy rate for the free report group in the first study (the moderate incentive condition). These results taken together indicate that metacognitive monitoring is effective at evaluating memory accuracy, and the control mechanism is able to screen out

answers that are more or less likely to be incorrect, depending on the context, by adjusting its response criterion.

A later study (Koriat & Goldsmith, 1996) further examined the role of monitoring effectiveness and control sensitivity in the trade-off between quantity and accuracy. In the first experiment of their study the researchers crossed test format (recall v. recognition) and incentive to be accurate (moderate v. high). Students answered general knowledge questions and were required to provide a response, and later a confidence rating, for each question. Finally the students decided if they wanted to provide or withhold their response for each question, based on one of two payoff matrices. The results replicated the researchers' earlier findings (Koriat & Goldsmith, 1994) that participants could screen out answers that were likely to be incorrect and thus increase their accuracy rates, and that this ability was under strategic control. The researchers also found a strong positive correlation between the subjective confidence in each response and the accuracy of the response, indicating effective monitoring. A strong positive correlation between confidence ratings and the likelihood that a response would be volunteered was also found, indicating high control sensitivity.

In the second experiment of their study, Koriat and Goldsmith (1996) used a similar procedure, and added some questions to which people are known to be highly confident in the wrong answer (deceptive questions; e.g., What is the capital of Australia? The answer is Canberra, not Sydney). They found that monitoring was ineffective for these questions and people were more overconfident in answers to deceptive questions than to non-deceptive questions. As the control process was based on the unreliable monitoring output participants were not able to reach high accuracy rates on the

deceptive questions. Essentially, the study showed that if monitoring is not effective, accuracy cannot be improved via providing a report option, regardless of control sensitivity. A similar study (Kelley & Sahakeyn, 2003) reported parallel results, and in addition found that when attention is divided at encoding the monitoring process suffers and the response criterion decreases. However, even when information was encoded under divided attention, participants could improve their accuracy rates by withholding responses; but they could not reach the same level of accuracy as those who encoded with full attention.

Clearly having the option to withhold a response allows for strategic control over accuracy rates through a quantity-accuracy trade-off. Yaniv and Foster (1995) proposed that control over precision, or grain size, also can improve accuracy via what they dubbed an informativeness-accuracy trade-off. (I will describe this trade-off as a "precision-accuracy trade-off". I use the term "informative" to describe the overall usefulness of responses; thus informativeness is a function of both the quantity and precision of responses.) Adjustment of response precision to increase accuracy is similar to withholding a response when one is not confident in its accuracy. However instead of withholding a response completely, a less precise response, in which one has more confidence (e.g., dark car as opposed to green car), is provided.

The Role of Control over Response Precision

Fisher (1996) reported that when participants provided free recall responses after viewing a filmed robbery neither their accuracy rates nor the number of propositions provided declined after several weeks. While there was no evidence of a quantity-accuracy trade-off, closer examination of the participants' statements revealed a

precision-accuracy trade-off. The stable accuracy rates were due to a decrease in the precision of the participants' statements as time passed. As their memories for coarse grained information lasted longer than their memories for more fine grained details, the participants modified their responses to be less precise, but correct, and therefore were able to maintain a consistent level of accuracy *and* quantity over time

Yaniv and Foster (1997) conducted three similar experiments examining the precision of responses; they asked participants to answer quantitative general knowledge questions (e.g., What was the date of the first trans-Atlantic flight?) in one of several ways. In the first study they provided the participants with scales differing in level of precision so that the participants could choose their own level of specificity (e.g., between 1700 and 1800, between 1700 and 1750, between 1715 and 1720), but required participants to answer every question. They told the participants to answer the questions as they felt the most comfortable, as if they were having a casual conversation with a friend. In the second study they asked the participants to provide a 95% confidence interval when responding, and the instructions stressed informativeness in responses. In the third study they asked participants to provide a point estimate and then the expected error as a plus or minus judgment (e.g. 1920, +/- 10 years).

The hit rates for all three studies were very low considering that the participants could always have chosen to provide very uninformative, but likely correct, responses. The first study had the highest accuracy rate of the three studies (55%). This was due to use of the scale with the largest grain size (which always included the correct answer); after excluding responses that used this scale the hit rate was only 46%, which was similar to the rates in the second and third studies (43% and 45% respectively). The

underutilization of coarse grained responses clearly shows that there is a strong desire to be precise and that people are willing to sacrifice accuracy to avoid vague answers. This need to be precise is consistent with both recent (Ackerman & Goldsmith, in press), and previous research (Bar-Hillel & Neter, 1993). Ackerman and Goldsmith found that when given the choice to sacrifice accuracy or precision, participants chose to forego accuracy. Similarly, Bar-Hillel and Neter found that people made excessively precise predictions, and did so even when providing accurate information was rewarded.

The drive to be informative is consistent with conversational norms, which dictate that we maintain both accuracy and informativeness (Grice, 1975). One reason for this need to be informative may be that rewards for informativeness are generally immediate whereas rewards for accuracy (or punishments for inaccuracies) are often delayed (Yaniv & Foster, 1997). When a request for information is met with an informative response, there is usually a reward for providing that answer, which may be as simple as a nod of thanks, or may be as significant as a word of approval from one's supervisor. This reward often comes regardless of the accuracy of the information supplied, because it takes time to determine if the information provided is in fact reliable. Likely due to this desire for informativeness, people tend to prefer responses that are precise, yet technically incorrect, to those that are technically correct, but imprecise (Yaniv & Foster, 1995). Therefore, when evaluating a response people take both absolute error and interval width into account. For example, a point estimate that is close to the correct answer, but incorrect, may be preferable to a wide range of possible answers which actually does include the correct response. (e.g., If the correct answer is 130 and the possible responses are 125, or, between 100 and 300, 125 may be the preferred response.)

Goldsmith, Koriat, and Weinberg-Eliezer (2002) used a different paradigm when examining precision, which more closely maps onto the studies on report option discussed above. In the first experiment of the study the participants answered quantitative general knowledge questions with a narrow interval and then provided a wider interval, with the interval widths set by the experimenter for each question. After providing the two intervals the participants indicated which of the intervals they would prefer to volunteer as their response, assuming they were "an expert witness testifying before a government committee". The participants provided the finer grained response 41% of the time and the coarser grained response 59% of the time, indicating they did not always choose to be either more precise or more accurate. The average accuracy rate of the volunteered responses was less than the rate that would have been achieved if a strategy of always choosing the wider interval had been used. However, the accuracy rate of the volunteered responses was higher than the rate that would have resulted from a strategy of always choosing the narrower interval. This provides more evidence of a balancing of, or trade-off between, accuracy and precision. In the second experiment the same procedure was used, but participants additionally provided confidence judgments for each of their answers. The results replicated those of the first experiment, and showed the participants were overconfident for both estimates, but more so for the narrower intervals. In the third experiment the procedure was the same as in the second, but participants received explicit payoff incentives. The results from experiments one and two replicated in study three.

The results of experiment three also showed that participants sacrificed accuracy more when the payoff for precision was higher, indicating that the participants could

strategically control the grain size of the information they provided in order to increase their earnings. This parallels the results of previous work (Koriat & Goldsmith, 1994), which found that participants could strategically control whether or not to withhold a response depending on the payoff matrix. The results of this experiment also revealed that people tend to provide the most precise answer they can that meets a preset response criterion (a "satisficing" model). The best predictor of whether the participants provided the more or less precise response was their confidence rating in the more precise response. More recent research (Ackerman & Goldsmith, in press) has expanded upon this model and is described below.

Using similar methodology Goldsmith, Koriat, and Pansky (2005) examined accuracy and precision over longer retention intervals. Their participants read a police interview that contained quantitative information and then answered questions about what they read. Participants provided one precise response (a single number) and then an interval response of a specified width. After providing the two answers the participants provided confidence judgments for both. The immediate test group then indicated which response they preferred to volunteer. The other two groups (one-day delay and one-week delay) returned on their specified day and indicated which responses they would volunteer. (The participants' original confidence judgments were not accessible to them when making these decisions about which response to volunteer.) The results indicated that the coarser estimates were more accurate than the finer estimates, and that as time passed the participants were more likely to volunteer the coarser response. The respective percentages of volunteered coarse responses for the immediate, one-day, and one-week conditions were 43%, 61%, and 75%. There was a drop in accuracy from the immediate

test condition to the one-day retention interval; however accuracy was maintained from the one-day retention interval to the one-week retention interval. Apparently, people can change their response strategies to maintain a level of accuracy as their memory for details declines.

Using Both Report Option and Control over Response Precision

A recent study expanded this line of research in several ways (Weber & Brewer, 2008). These researchers were interested in determining if the findings regarding precision would generalize to eyewitnesses, as previous studies had used only general knowledge tests and memory for written material to examine strategic control over precision. Therefore Weber and Brewer used stimulus material that was more ecologically valid than materials used in previous studies; specifically their participants watched a videotaped mock bank robbery. The researchers also increased the ecological validity of their design by including questions addressing both quantitative and non-quantitative details (specifically, color information). In addition, they incorporated both control over precision and response option into one study, which had not been done in any previous research. They found that their participants demonstrated an accuracy-precision trade-off and that when given the option to provide a fine-grained, coarse-grained, or no response, the eyewitnesses used both grain size adjustments and the report option to balance their accuracy and informativeness. This is important because in the real world there is almost always the possibility of either adjusting the grain size of a response, or of not responding at all. It appears that people employ both of these techniques to maintain an acceptable level of accuracy.

Recently Ackerman and Goldsmith (in press) put forward a model that accounts for decisions to withhold a response as well as decisions to adjust precision. Their dual-criterion model is an extension of the satisficing model proposed by Goldsmith, Koriat, and Weinberg-Eliezer (2002). According to the original satisficing model the response provided is the most precise response possible that meets a minimum confidence threshold. Ackerman and Goldsmith's model suggests that people choose to provide responses that meet both a minimum confidence criterion *and* a minimum precision criterion. According to this model, if people have knowledge that meets both the confidence and precision criteria (i.e., satisficing knowledge) then they will behave in a way consistent with the satisficing model and provide the most precise response that meets their confidence criterion.

The difference between the satisficing model and the dual-criterion model lies in how someone without satisficing knowledge behaves. The satisficing model predicts that people with unsatisficing knowledge will provide a very coarse response in which they are sufficiently confident. In contrast, the dual-criterion model predicts that, rather than provide a response that is extremely imprecise, someone with unsatisficing knowledge will, given the opportunity, choose to say "I don't know" rather than provide a response that is either excessively broad, or likely to be incorrect. It is also expected that without a report option people with unsatisficing knowledge will choose to violate the confidence criterion instead of the precision criterion. Supporting their theory, Ackerman and Goldsmith (in press) did find that, given an "I don't know" option individuals with unsatisficing knowledge exercised their report option, however without an "I don't know" option, participants chose to sacrifice accuracy rather than precision.

The Current Study

As the studies reviewed above indicate, both report option and control over precision are effective ways to maintain acceptable accuracy rates. The current study took several steps to advance this research area further. The first was to manipulate the emphasis on accuracy versus informativeness. Previous research either did not manipulate the incentive to be accurate versus informative or, provided monetary incentives to be accurate. In the current study to make the emphasis on informativeness or accuracy more applicable to eyewitnesses, it was conveyed during the interaction between the interviewer and the witness. Witnesses know they are expected to be informative, but at the same time should be aware that a mistake on their part could have significant consequences. The information they are provided regarding the goals of an interview (i.e., "we need as much information as possible" vs. "we need accurate information") may be important when setting a response criterion. Thus, I expect accuracy instructions to increase the responses criterion, making it more difficult to meet. This should lead to a decrease in the amount of information provided, a decrease in the precision of the information provided, an increase in the number of "I don't know" responses, and an increase in the accuracy rate of the information. In contrast, the informativeness instructions should lower the criterion, allowing additional information to be provided while lowering the accuracy rate. Informativeness instructions should also increase the precision of responses, as relatively more precise responses should be able to meet the lowered response criterion.

The second step was creating a more ecologically valid paradigm for examining the effects of metacognition on eyewitnesses' reports. Most of the researchers cited have

mentioned eyewitness testimony as a relevant example of metacognition at work. However, until Weber and Brewer (2008) used a video of a mock crime as their stimulus material, and Goldsmith, et al. (2005) used a transcript of an interview with an eyewitness, no study had actually used a paradigm that was related to episodic memory in any way; most had used general knowledge tests. It could be argued that general knowledge tests are not testing memory in some cases, as there is no way to verify that the participants ever knew the answers to the questions. In addition to using ecologically valid stimulus materials, an ecologically valid interview procedure was employed. Prior studies have tested memory in a written test format, and have often required participants to provide either (a) a response to every question regardless of whether they knew the answer, and then determine if they would volunteer that response (e.g., Koriat & Goldsmith, 1996) or (b) multiple responses at different levels of precision, and then indicate their preferred answer (e.g., Weber & Brewer). In actual interviews witnesses are interviewed orally; they are not asked to fill out a form or provide a written account. Nor are they required to generate multiple responses and then choose which one to volunteer. Using an oral interview allows for a superior assessment how an actual eyewitness would use metacognition to maintain accuracy. However, this oral interview will be followed by a forced response phase in which the participants will be asked to pretend that a report option had not been available, and provided confidence judgments for all of their responses. I anticipate that, in line with prior research (e.g., Koriat & Goldsmith, 1994; Weber & Brewer, 2008) confidence ratings for volunteered responses will be higher than confidence ratings for responses that were only provided when "I don't know'" was no

longer an option. This is because participants will be reasonably successful at monitoring their confidence and filtering out any responses that are unlikely to be correct.

The last step was to continue the line of inquiry started by Weber and Brewer (2008) by examining the combined effects of report option and control over precision on memory accuracy rates for both quantitative and non-quantitative information, and to do so for both immediate and delayed recall. I expect accuracy rates to be maintained across the retention interval because the participants will exercise metacognitive control in whatever way they are able. However this metacognitive control will lead to trade-offs between accuracy and both precision and quantity as time passes. Therefore, delayed reports will be less precise and/or include fewer pieces of information than immediate reports. For cued and yes/no question formats I anticipate more "I don't know" responses at delayed recall than at immediate recall, further demonstrating use of the report option.

It is important to consider both report option and control over response precision in the same study, as both mechanisms can be employed in the real world, and both are used according to recent findings (Ackerman & Goldsmith, in press; Weber & Brewer). In the current study, question type was varied in order to look at these two metacognitive mechanisms in a novel, more ecologically valid way. In a yes/no question condition the level of precision and the maximum quantity of information were controlled by the experimenter, as the participants could only say yes, no, or "I don't know" to each question as asked. Thus, I expected the highest number of "I don't know" responses in the yes/no condition, as report options was the only viable mechanism. In a cued question condition the maximum quantity of information was somewhat constrained by the number of questions asked, but precision was under the participants' control. Finally,

both the quantity and precision of information were controlled solely by the participants in the free narrative condition. Due to this optimal level of control, I expected the free narrative condition to yield the highest accuracy rate. This method of controlling precision and quantity is more in line with real world interviewing procedures than methods used in prior studies, such as requiring participants to provide estimates at a predetermined level of precision.

CHAPTER II

Method

Design

The study employed a 2 (retention interval: immediate v. one week delay) x 3 (question format: free narrative v. cued v. yes/no) x 2 (interviewer instructions: be accurate v. be informative) between subjects factorial design. Participants watched a video of a mock crime and were then interviewed about their memory of the crime. Retention interval was manipulated such that half of the participants were interviewed immediately, and the other half returned after one week for their interview.

The participants were questioned about the crime they viewed in one of three styles (See Appendices A, B, and C for the interview questions). They were either asked to provide a free narrative, were asked a series of cued questions (e.g., "Please describe, if you can, the thief's hair."), or were asked a series of yes/no questions (e.g., "Did the thief have brown hair?"). Prior to the questioning, half of the participants were told that they were the only witness, and in order for the crime to be solved they must provide as much information as possible (informativeness instruction condition). The rest of the participants were told that the detective had limited resources to follow up on any information provided, so the witness should only provide very accurate information (accuracy instruction condition; example instructions are provided in Appendix D).

To facilitate the creation of the interviews, the video of the mock crime was shown to approximately 50 students in an introductory psychology class in exchange for extra credit. Participants were immediately asked to provide a written free recall of everything they remembered about the crime. They were asked to provide as much detail

as possible, without placing much weight on accuracy. Their responses were used to create a list of the information likely to be supplied by participants in the actual study (e.g., the thief's hair color, what was stolen) and to create the cued and yes/no questions for the actual study. Questions were not limited to information commonly reported during the pilot to avoid ceiling effects. Facts reported at various frequencies were used to create the 15 question interview.

The yes/no interview was composed of some questions for which the accurate response was "yes" and some questions for which the appropriate response was "no". For example, in the crime video the thief wore a denim jacket and jeans, so questions such as "Was the perpetrator wearing khakis?" (false) and "Was the perpetrator wearing a denim jacket?" (true) were used in the yes/no interview. Four counterbalanced versions (see Appendix A for an example) of the yes/no interview were used so that for certain items the correct answer could be yes, or no, depending on the interview version. Using the shirt and pants example above, another version of the interview would include the questions "Was the perpetrator wearing jeans?" (true) and "Was the perpetrator wearing a blazer?" (false).

All interviews were videotaped, transcribed, and scored. The dependant variables include accuracy, confidence ratings, quantity of correct information, number of "I don't know" responses, and precision. The scoring system is described below.

Participants

One hundred forty-nine students at Florida International University were recruited to participate in a study on "Perceptions of Actors in Comedies and Dramas" from the psychology department participant pool via Sona-Systems. Students signed up for two

appointments one week apart. They were awarded two research credits for their participation, to be applied towards a psychology course research requirement. Each time slot was assigned randomly to an experimental condition before the participant arrived at the lab. Participants in the immediate retention interval condition were not required to return for their second appointment.

The final sample consisted of 126 students (the data elimination procedures are described in detail in the results section). The average age was 20 years, and the majority of the participants were female (66%; 34% male) and Hispanic (55%; 17% White, 14% African-American, 10% Asian, 4% other).

Procedure

Upon arrival at the lab participants were told that they were about to take part in a study in which they would watch a short video and answer questions about what they saw. The participants were not told that a memory test would follow the video, in order to prevent rehearsal in both the immediate and delayed recall groups. The participants then watched a 50 second video of a mock crime in which a man steals a bag from a woman and gets away on a bicycle. Participants then completed a five minute filler task consisting of basic arithmetic problems. Next they provided demographic information and answered several questions about the quality of the acting, directing, and production of the crime video (e.g., What did you think about the production quality of the video?). These irrelevant questions were designed to ensure that participants believed the purpose of the study corresponded with the title of the study, “Perceptions of Actors in Comedies and Dramas”, and did not rehearse over the delay.

At this point participants assigned to the delayed recall group were allowed to leave, and reminded to return in one week for their follow up session; participants in the immediate recall condition began their interview and were given instructions, and asked questions, corresponding with the condition (See Appendices A, B, and C for the interview questions). Regardless of the questioning style, all participants were instructed that they had a report option (See Appendix D for example instructions). In the free narrative condition participants were told to "Keep in mind that if you don't remember something you should not guess, instead just leave out the information". In the cued and yes/no questions conditions participants were told "Keep in mind that if you don't remember something you should not guess, instead just say "I don't know" or "I don't remember"". During this instruction phase all participants were told to assume that the interviewer was the detective assigned to the case. The entire interview was recorded, transcribed, and scored. After the interview was over, the participants in the cued and yes/no conditions were asked to look over the answers they provided (i.e., all but the "I don't know" responses) and write down their confidence for each one on a scale from 0 – 100. The participants provided these judgments on the piece of paper listing the questions, which included the interviewer's hand written transcription of their responses. They were then asked to go back to the questions they chose not to answer and write down what they would have said if they had been required to provide a response, and write down their confidence in that new response. The delayed recall participants returned to the lab one week later and underwent the same interview process as the immediate interview group.

Scoring

The main dependent variables of interest are: quantity of information, accuracy of information, precision of information, and number of "I don't know" responses. Descriptions of how each of these was scored are provided below. The basic structure of the scoring system was to score the response to each question in one row (see Appendix). For example, the first question dealt with the thief's age. So, any information provided in responses to this question would be scored in the columns of first row, labeled "Q1". The second question dealt with the victim's age, and any information provided in response to this question was scored in the second row, labeled "Q2." This structure was quite straightforward for the cued and yes/no conditions. However for the free narrative condition the process was slightly different, as in this condition there were not 15 questions, but instead only one.

For the free narrative condition the scorers scored the information provided in whichever row best corresponded with the content of the response. Thus, any information about the thief's age would be recorded in the Q1 row, and any information provided about the victim's age would be recorded in the Q2 row. If information was provided that did not correspond with any of the 15 questions, the scorers recorded that information in the extra rows provided, labeled "other 1", "other 2", etc. The most common "other" responses were: information about the setting (e.g., it was on a college campus), information about the victim not covered by the interview questions (e.g., the victim was African-American), and information about the other people in the video (e.g., there were other people around who saw what happened). While this system worked well, it does mean that the scorers did not always break up the narrative in the exact same way and

one scorer might fill in more rows than another. However, as the overall scores were of interest, this was of minimal importance.

All of the transcripts from the yes/no condition were scored by the PI, as the scoring of these was objective once the scoring system was in place. The PI was blind to all conditions other than interview type, as were all scorers. The scorers were trained for approximately six weeks prior to beginning their task, and had meetings with the PI approximately weekly while scoring the transcripts. Of the 82 cued and free narrative transcripts, 26 (32%) were scored by two scorers; the rest were scored by only one. For these 26 transcripts, inter-rater reliability was determined by looking at (a) the correlation of the scores between the two scorers, and (b) a t-test to verify the means of the two scores were not significantly different from each other. After discussions between the scorers and the PI, double checking for errors in all scoring sheets, and having the PI double check long and/or difficult transcripts, the inter-rater reliability was high. The results of the reliability analysis are reported below along with the description of each dependent variable.

Scoring quantity. The general policy on scoring for quantity was that any noun (e.g., jeans), adjective (e.g., old), active verb (e.g., grabbed), and adverb (e.g., slowly) would count as one piece of information. Prepositions that clearly defined a relationship were also counted (e.g., he rode towards her). For example, if someone reported "She was walking quickly towards a fence and wearing a long brown coat", this would be seven pieces of information (walking, quickly, towards, fence, long, brown, coat). Or, if someone reported "He grabbed her bag and rode a bike away slowly" this would be five pieces on information (grabbed, bag, rode, bike, slowly).

A piece of information was counted only once if it was repeated, unless it was repeated in a different context. For example, if the participant said "she was carrying a bag" and later said "he stole her bag", "bag" would be counted twice. However, if the participant said "she was carrying a bag on her arm. It was a big black bag…" then "bag" would only be counted once. When something was negated this was counted as one piece of information. Thus, "He didn't run away very fast" would be one piece of information. If multiple pieces of information were provided that differed in level of specificity, only the most precise was counted and scored for accuracy. For example, once the information "beard" had been provided, "facial hair" did not provide more information.

The total quantity measure was the sum of the quantity scores across all the rows. The correlation between the two scorers total quantity scores was very high ($r = .98$, $p < .001$) and there was no significant difference between the means ($p = .807$).

Scoring accuracy. The next step was to evaluate the accuracy of each piece of information recorded in the "Quantity" column. Each piece could be scored as either accurate, inaccurate, or unscoreable. To determine what responses should be considered accurate, 13 research assistants were shown the video and asked to answer the questions from both the cued and yes/no interviews. They were not asked to *remember*, but to answer the questions while watching the video, and to go back and watch the video again if necessary. This was to provide an idea of what should be considered an "accurate" response for questions in which the "accurate" response is not immediately apparent (e.g., the age of the thief and victim). If a participant said something that was not listed as either accurate or inaccurate based on the research assistants' responses, the scorers checked the video themselves. If the scorers were unsure of the accuracy after watching

the video the information was either discussed with the PI, or scored as "unscoreable". Information was recorded as unscoreable if it was unverifiable (e.g., "it was late winter"). If the participant described something that was not on the video but could have been true (e.g., "the victim was wearing a white shirt", when you couldn't see her shirt under the jacket) this was counted as unscoreable. However, if the participant described something that was not on the video and could not have been true (e.g., "the thief was carrying a baseball bat", which he was not) this was scored as inaccurate. The accuracy rate was computed by dividing the total number of accurate pieces of information by the total quantity (after deducting the unscoreable information from the total quantity). The inter-rater correlations for number of accurate pieces of information, number of inaccurate pieces of information, number of unscoreable pieces of information, and accuracy rate were all high (rs = .98, .96, .94, and .94 respectively, all ps < .001). There were no significant differences between the means on any of the accuracy measures (all ps > .250).

Scoring precision. Precision was rated on a five point scale, with four and five being high, and one and two being low. A three was considered a "base" response, or a response that was more or less typical. It was stressed to all scorers that precise, yet incorrect, information should be scored the same as equally precise correct information. Scorers were provided with a list created by the PI of common responses and their respective scores. For example, for the question "Please describe the thief's pants", the "base" response was "jeans", and thus was scored as a three. "Blue jeans" was also scored as three, as blue is the default and providing that information does not help to identify the jeans worn beyond the information "jeans". However, "dark blue jeans"

would be scored as a 4, and "dark blue jeans with a belt" would be scored as a five. Something less specific than "jeans", such as "pants", would be scored as a one. As participants did not always provide a response that was on the list given to the scorers, the scorers either retrieved the precision score directly from the list, or evaluated the precision themselves based on the examples on the list.

The sum of the precision scores across all of the questions provided a "total precision" measure. This was then divided by the number of rows filled in, to create an "average precision" measure. The inter-rater correlations for total precision, and average precision were both satisfactory (rs = .96 and .74, respectively, all ps < .001). There were no significant differences between the means (all ps > .700)

Scoring "I don't know". When a participant said "I don't know" or "I don't remember" to a question, the "IDK" column was checked. The IDK column was only checked when absolutely no information was provided in response to the question. So, a response of "I don't know; he was wearing jeans I guess" would be scored as one piece of information (jeans), not as IDK. This column was also checked if the participant provided a response and then retracted it. The inter-rater correlation for the number IDK responses was high (r = .97, p < .001) and there was no significant difference between the means (p = .336).

Chapter III

Results

Data Elimination Procedures.

Of the 149 participants who completed this study, the data from 126 are included in the sample. The data from eight participants were excluded because of errors by the research assistant (e.g., forgetting to record the interview, recording the interview with no sound, and explaining the task incorrectly). One participant misunderstood the task and was eliminated, and another was unwilling to follow the instructions and was eliminated. Another participant was deaf and therefore responded to the interview in a written format instead of orally, and thus his data were not included. One outlier was identified and eliminated from the analyses. As recommended by McClelland (2000), any changes to the conclusions of the analyses that result from keeping the outlier in the sample will be mentioned in the text. There was only one difference to report, and it is described in the "Total quantity" section below.

The data collected from 11 participants were deleted because of changes to the study materials. When data collection began two versions of the cued interview were included. The first version comprised questions which were likely to obtain one word answers (e.g., What kind of pants was the thief wearing?). The second comprised questions which gave the participants the opportunity to elaborate more (e.g., Please describe, if you can, the pants the thief was wearing.). After running several participants it was determined that the second version would be used, as it appeared to motivate more elaborative responses. Thus, the five people who were interviewed with the first cued version were eliminated. Similarly, after running several participants in the free narrative

condition it became apparent that participants were not providing many descriptive details. Therefore, the instructions were modified to specifically request descriptive details in addition to actions. Thus, six people were eliminated because they received the original instructions.

Instruction Manipulation

The instruction manipulation, which stressed either accuracy or informativeness, was unsuccessful. There was no effect of instructions on the quantity of information provided ($F(1,114) = 1.85$, $p = .177$). The instruction manipulation did not interact with either interview type ($F(2,114) = 1.23$, $p = .297$) or retention interval ($F(2,114) = 0.01$, $p = .918$) to affect quantity of information. There as also no three-way interaction between the independent variables on quantity ($F(2,114) = 1.29$, $p = .280$).

There was also no effect of the instructions on the accuracy rate of the information provided ($F(1,114) = 1.34$, $p = .249$). There was no interaction between instructions and interview type ($F(2,114) = 0.65$, $p = .526$), nor was there an interaction between instructions and retention interval ($F(1,114) = 2.56$, $p = .112$), on accuracy rate. The three-way interaction on accuracy rate was also not significant ($F(2,114) = 1.60$, $p = .206$).

There was no effect of instructions on the number of "I don't know" responses provided ($F(1,77) = 0.01$, $p = .919$). There was no interaction on number of "I don't know" responses between instructions and interview type ($F(1,77) = 0.33$, $p = .570$), nor was there an interaction between instructions and retention interval ($F(1,77) = 0.20$, $p = .653$). The there-way interaction was also non-significant ($F(1,77) = 1.96$, $p = .165$).

Finally, instructions did not affect average precision of responses ($F(1,114) = 0.02$, $p = .903$). There was no interaction between instructions and interview type ($F(2,114) = 0.003$, $p = .997$), nor was there an interaction between instructions and retention interval ($F(1,114) = 0.10$, $p = .748$), on average precision. The three-way interaction on average precision was also not significant ($F(2,114) = 0.02$, $p = .982$). As the instruction manipulation had no effects on any of the dependent variables it will not be discussed further.

Quantity

Total quantity. The measure of total quantity of information provided includes information that was subsequently scored as accurate, inaccurate, and unscorable. (The quantity of accurate information only is analyzed below.) There was a main effect of interview type ($F(2,120) = 31.93$, $p < .001$, partial $\eta^2 = .347$, power = 1.00) on total quantity. Post-hoc comparisons showed that participants in the cued (M = 38.36) and free narrative (M = 34.63) conditions provided more information than the people in the Y/N condition (M = 19.12; both ps < .001). The difference between the cued and free narrative conditions was not significant ($p = .212$). The differences remain significant after applying a Holm-modified Bonferroni (HMB) adjustment to correct for multiple comparisons. It should be noted that this effect is somewhat arbitrary as the number of questions and wording of the questions asked in the Y/N and cued conditions determined the maximum quantity of information than could be provided.

There was also a main effect of retention interval ($F(1,120) = 9.91$, $p = .002$, partial $\eta^2 = .076$, power = .878) on total quantity: participants who were interviewed immediately provided more information (M = 33.78) than those who were interviewed

after a week long delay (M = 27.53). There was no interaction between the retention interval and the interview style on total quantity ($F(2,120) = 2.24$, $p = .111$, partial $\eta^2 = .036$, power = .45). If the data from the outlier had been included in the analysis, this interaction would have been significant ($p = .048$). As the direct comparisons of quantity between the different interview conditions are not meaningful, this is not considered an important inconsistency. However, to further explore this potential interaction, tests of the simple main effects (see Table 1 for cell means) were conducted to determine if there was an effect of retention interval in any of the conditions. There was no effect in the Y/N condition ($t(41) = 1.01$, $p = .319$, $d = .32$), or the cued condition ($t(40) = 1.54$, $p = .133$, $d = .49$), but there was a significant decrease in quantity across the retention interval in the free narrative condition ($t(39) = 2.69$, $p = .010$, $d = .86$). This significant effect for the free narrative condition remains significant after correcting for multiple comparisons using the HMB.

Quantity of accurate pieces of information. In addition to measuring total quantity, the quantity of correct information provided was also examined, as this is the measure of "memory" generally utilized by the quantity-oriented approach to memory. Using this measure of quantity the pattern of results is very similar to the pattern of results found for the total quantity measure. Specifically, there is a main effect of interview type on the quantity of accurate information ($F(2,120) = 32.13$, $p < .001$, partial $\eta^2 = .349$, power = 1.00). Post-hoc comparisons revealed that just as with total quantity, the quantity of accurate information in the Y/N condition (M = 17.19) was less than in both the cued condition (M = 33.43) and the free narrative condition (M = 32.17; both ps $< .001$). Once again the difference between the cued and free narrative conditions was not

significant (p = .731). After applying the HMB to correct for multiple comparisons these effects remain significant. It should be noted again that these effects are somewhat arbitrary as the number of questions, and the wording of the questions asked in the Y/N and cued conditions determined the quantity of information than could be provided. There was also a main effect of delay on the quantity of accurate information (F(1,120) = 13.97, p < .001, partial η^2 = .104, power = .96). Participants interviewed immediately provided more information (M = 30.79) than participants interviewed after a one-week delay (M = 24.27).

The only difference between the pattern of results for the quantity of accurate information and for total quantity is that for the quantity of accurate information there was a significant interaction between retention interval and interview type (F(2, 120) = 3.09, p = .049, partial η^2 = .049, power = .59; Please note this interaction was significant for the total quantity measure when the outlier was included in the analysis). As the quantity of information in the Y/N condition was somewhat predetermined, direct comparisons between interview conditions are not very meaningful. Thus, to further explore this interaction tests of the simple main effects were conducted. These tests showed (see Table 1 for cell means) that there was no effect of retention interval in the Y/N condition (t(41) = 1.06, p = .294, d = .33), a marginal effect in the cued condition (t (40) = 1.88, p = .067, d = .60), and a significant effect in the free narrative condition (t(39) = 3.15, p = .003, d = 1.01). The significant effect in the free narrative condition remains significant after applying the HMB correction for multiple comparisons.

Accuracy

Accuracy rate of volunteered responses. This measure of accuracy reflects the accuracy rate of the responses provided freely during the interview, and does not include the accuracy rate of the responses given in the post-hoc forced response part of the study. There was a main effect of interview type on accuracy rate ($F(2,120) = 8.98$, $p < .001$, partial $\eta^2 = .130$, power = .97). Post-hoc tests revealed that the accuracy rate in the free narrative condition (M = 94.29) was higher than the accuracy rate in the Y/N (M = 89.85; $p = .004$) and cued conditions (M = 87.97; $p < .001$). These differences remain significant after correcting for multiple comparisons using the HMB. The difference in accuracy rate between the cued and Y/N conditions was not significant ($p = .224$). There was only a marginal effect of delay on accuracy rate, ($F(1,120) = 3.31$, $p = .072$, partial $\eta^2 = .027$, power = .44) such that the accuracy rate was higher at the immediate interview (M = 91.73) than the delayed interview (M = 89.64). There was no interaction between retention interval and interview style on accuracy rate ($F(2,120) = 0.30$, $p = .745$, partial $\eta^2 = .005$, power = .096). It should be noted that the observed power to detect an effect of the delay was very low, as was the power to find an interaction. Simple main effects (see Table 1 for cell means) were examined to determine if there was a significant decrease in accuracy rate across the retention interval for any of the interview conditions. These analyses showed that there was no change in accuracy rate in the Y/N condition ($t(41) = 0.45$, $p = .652$, $d = .14$), the cued condition ($t(40) = 1.43$, $p = .160$, $d = .45$) or the free narrative condition ($t(39) = 1.43$, $p = .162$, $d = .46$).

Complete accuracy rate. Another analysis was done on the "complete" accuracy rate, which includes both volunteered responses, and responses provided in the post-hoc

forced response phase of the study, to determine whether the accuracy rate would have declined in the Y/N and cued conditions without the option to say "I don't know". This accuracy rate is the percentage of the 15 questions that were answered correctly when both volunteered and withheld responses are considered, and is therefore not directly comparable to the accuracy rate discussed above which scored by piece of information, not overall response. There was a main effect of delay on the "complete" accuracy rate, ($t(75) = 2.27$, $p = .026$, $d = .52$). The "complete" accuracy rate was higher at the immediate interview (M = 82.70) than at the delayed interview (M = 76.17). This decrease across the delay indicates the null finding using the accuracy rate of volunteered responses only was not due to a ceiling effect because of an insufficient delay.

I Don't Know Responses

Quantity of "I don't know" responses. There were main effects of both interview type ($F(1,81) = 8.88$, $p = .004$, partial $\eta^2 = .099$, power = .837) and retention interval ($F(1,81) = 6.22$, $p = .015$, partial $\eta^2 = .071$, power = .693). Participants in the free narrative condition were not included in this analysis as they did not have an opportunity to say "I don't know". Participants in the Y/N condition said "I don't know" more (M = 2.77) than participants in the cued condition (M = 1.67) and participants who were interviewed after a one-week delay said "I don't know" more often (M = 2.69) than those who were immediately interviewed (M = 1.77). There was no interaction between retention interval and interview type on the number "I don't know" responses provided ($F(1,81) = 0.51$, $p = .479$, partial $\eta^2 = .006$, power = .11). Once again, the low power to find an interaction should be noted.

Confidence in "I don't know" responses. Among the participants who said "I don't know" at least once (n = 70), the average confidence in volunteered responses was higher (M = 88.52) than the average confidence in the originally withheld responses (M = 48.18; $t(69) = 18.25$, $p < .001$, $d = 4.39$). Among all the participants in the Y/N and cued conditions, regardless of whether they provided any "I don't know" responses, the average confidence across all 15 questions (M = 82.68) was significantly lower than confidence in their volunteered responses (M = 88.71; $t(84) = -9.67$, $p < .001$, $d = -2.11$).

"I don't know" and accuracy. For the Y/N and cued conditions, among participants who said "I don't know" at least once, the accuracy rate of the responses that were volunteered (M = 85.21) was significantly higher than the accuracy rate of the responses that were initially withheld (M = 47.87; $t(64) = 7.99$, $p < .001$, $d = 2.00$). However, for the Y/N and cued conditions, there was no correlation between saying "I don't know" more often and accuracy rate ($r = .017$, $p = .878$).

Precision

There were main effects of both interview type ($F(2,120) = 281.04$, $p < .001$, partial $\eta^2 = .824$, power = 1.00) and delay ($F(1,120) = 6.33$, $p = .013$, partial $\eta^2 = .050$, power = .70) on average precision. The average precision of cued responses (M = 3.49) and free narrative responses (M = 3.33) was higher than the average precision of Y/N responses (M = 1.91; both ps < .001). The average precision of cued responses was also significantly higher than the average precision of the free narrative responses ($p = .046$). All of these remain significant after correcting for multiple comparisons using the HMB. The average precision was higher for immediate responses (M = 2.96) than delayed (M = 2.84). There was also a marginally significant interaction ($F(2,120) = 2.77$, $p = .066$;

partial η^2 = .044, power = .54). As the precision of responses in the Y/N condition was predetermined, comparisons between the Y/N interview conditions and the other interview styles are not meaningful. Thus, to explore this interaction an analysis of the simple main effects was conducted. This analysis showed (see Table 1 for cell means) that there was a significant decrease across the delay in average precision in the cued condition ($t(40) = 2.20$, $p = .033$, $d = .70$) and a marginal decrease in the free narrative condition ($t(35.50) = 1.95$, $p = .059$, $d = .65$)[1], but no effect in the Y/N condition ($t(41) = -0.91$, $p = .369$, $d = -.28$). After correcting for multiple comparisons using the HMB, the effect in the cued condition is no longer significant.

[1] Levene's test for inequality of variance was significant for this comparison, thus the results reported do not assume equal variances.

Table 1.

Cell Means

Interview	Interval	Total Quantity			Total Accurate			Accuracy Rate			I Don't Know			Avg Precision		
		Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n
Yes/No	Immediate	19.63	2.72	22	17.73	2.73	22	90.38	7.11	22	2.18	1.30	22	1.89	0.13	22
	Delay	18.57	4.10	21	16.62	4.02	21	89.30	8.41	21	3.38	2.42	21	1.94	0.21	21
	Total	**19.12**	**3.46**	**43**	**17.19**	**3.42**	**43**	**89.85**	**7.70**	**43**	**2.77**	**2.00**	**43**	**1.91**	**0.17**	**43**
Cued	Immediate	41.81	13.61	21	37.14	12.27	21	89.68	7.42	21	1.33	0.86	21	3.62	0.40	21
	Delay	34.90	15.48	21	29.71	13.28	21	86.26	8.07	21	2.00	1.92	21	3.36	0.35	21
	Total	**38.36**	**14.81**	**42**	**33.43**	**13.18**	**42**	**87.97**	**7.85**	**42**	**1.67**	**1.51**	**42**	**3.49**	**0.39**	**42**
Narrative	Immediate	41.11	16.59	19	38.89	15.18	19	95.56	4.05	19				3.46	0.30	19
	Delay	29.05	11.98	22	26.36	10.13	22	93.19	6.17	22				3.22	0.49	22
	Total	**34.63**	**15.37**	**41**	**32.17**	**14.05**	**41**	**94.29**	**5.37**	**41**				**3.33**	**0.43**	**41**
Total	**Immediate**	**33.73**	**15.99**	**62**	**30.79**	**14.69**	**62**	**91.73**	**6.85**	**62**	**1.77**	**1.17**	**43**	**2.96**	**0.85**	**62**
	Delay	**27.53**	**13.23**	**64**	**24.27**	**11.24**	**64**	**89.64**	**8.01**	**64**	**2.69**	**2.27**	**42**	**2.84**	**0.74**	**64**
	Total	**30.58**	**14.92**	**126**	**27.48**	**13.40**	**126**	**90.67**	**7.51**	**126**	**2.22**	**1.85**	**85**	**2.90**	**0.80**	**126**

Chapter IV

Discussion

Metacognitive Mechanisms for Maintaining Accuracy

Common belief is that memory declines with the passage of time, and thus reports given after a retention interval will be inferior to those given immediately. This inferiority is expected across a variety of measures, most notably, accuracy. That is, people expect memory reports to become less accurate with the passage of time. Yet in this study participants reached relatively high accuracy rates for all retention intervals. How can stable accuracy rates over time be reconciled with a declining memory? The current study suggests that there are two mechanisms which allow for a given accuracy rate to be reached and then maintained as time passes: withholding information and adjusting precision. Witnesses' use of these two mechanisms has implications for theories of metacognition, the criminal justice system, and researchers who study witness memory.

The first mechanism used to attain a desired accuracy level is to withhold information. By withholding information that the monitoring mechanism assesses as likely to be incorrect (or similarly, by saying "I don't know" when unsure of a response), the chances of providing incorrect information decrease. The current study provides evidence that this mechanism is used. The majority of participants in both the immediate and delayed conditions chose to say "I don't know" to at least one question. This suggests that regardless of when a memory test occurs, individuals are monitoring their memories and controlling what they report in order to achieve an acceptable accuracy rate.

It was possible to determine if the withheld responses actually were more likely to be incorrect than responses that were provided voluntarily because the design used in this experiment included a post-hoc forced response phase. During this phase participants provided responses to the questions which they said "I don't know" to initially. As expected, the answers participants chose not to provide when given a report option were less accurate, and were held with less confidence, than those volunteered. This means that the participants were able to monitor their knowledge and effectively determine whether they could provide a response that was sufficiently likely to be correct, and answer accordingly.

Effective monitoring allows for a relatively high accuracy rate to be attained through the withholding of certain information. However, this withholding of information implies that the price of maintaining accuracy is a decrease in the quantity of information provided. Just such a quantity-accuracy trade-off has been found in previous research (e.g. Kelley & Sahakyen, 2003; Koriat & Goldsmith, 1994, 1996). In order to examine this trade-off in the current study, both changes in quantity and accuracy over time were examined. Quantity of information provided decreased after a delay; by comparison, the accuracy rate did not decrease after a delay. In order to verify that this was in fact a "trade-off" over time (and not just a decrease in quantity), it had to be established that there would have been a decrease in accuracy over time had there not been an option to say "I don't know" (a report option). Otherwise it could be argued that the accuracy rate simply did not decline across the delay because the delay was insufficient to weaken the memory trace, or because the questions were too easy and thus accuracy rate was an insensitive measure. The data from the post-hoc forced report phase allowed us to

determine if the accuracy rate would have declined over time without a report option. When the forced report responses provided after the initial interview (i.e., the withheld responses) were combined with the freely provided responses to create a complete, or overall, accuracy rate, there was a decrease in accuracy after a delay. This indicates a decline in memory over time: participants were less able to answer accurately all the questions after a delay. Yet, this decrement was not reflected in their accuracy rates because they successfully utilized metacognitive monitoring and control to strategically withhold low-confidence responses.

The second mechanism allowing accuracy to be maintained is to adjust the precision of responses to be more or less specific. For example, when asked about the thief's clothing, participants could simply provide a vague response (e.g., "He was wearing dark clothes") which they are confident is correct, rather than attempt to provide a more precise response (e.g., "He was wearing a black raincoat"), which they are uncertain about. One way to examine the role of control over precision is by comparing how participants behaved in the different interview types, because the different interviews allowed for varying levels of control over precision. In the yes/no condition the participants had no control over the precision of their responses; they only had the options of answering the question as asked or saying "I don't know". In contrast participants in the cued condition could either say "I don't know" or adjust the precision of their responses. When control over precision is not possible, participants should rely more heavily on the first mechanism (report option) to reach an acceptable accuracy rate. This would lead to more "I don't know" responses in the yes/no condition than in the cued condition. This is the pattern found in the data: in the yes/no interview condition

participants said "I don't know" about once more on average than participants in the cued condition. Both groups were able to reach the same level of accuracy despite different levels of control over precision.

Another way to examine the role of adjusting precision to maintain accuracy is to compare the levels of precision chosen over time. As a memory trace fades, details may become less accessible, and thus precision may have to be sacrificed in order to reach an acceptable level of accuracy. This precision-accuracy trade-off pattern was found in the current study: the average precision of responses was lower at the delayed interview than at the immediate interview, while the accuracy rate was not. This supports the idea that as a memory trace fades, the precision of a response can be adjusted to maintain accuracy. Thus, using the "dark clothes" versus "black raincoat" example, both responses are correct, and the same quantity of information is provided. However, precision was sacrificed if the response "dark clothes" was provided[2]. The current study indicates that both the ability to withhold information and control over response precision are important mechanisms allowing people to maintain consistent accuracy rates. However, when only one of the two options is available, people can rely more heavily on the available mechanism without sacrificing accuracy.

Implications of Effective Metacognition in Witnesses

Errors of commission and errors of omission. Any concerns of attorneys, jurors, or investigators that witnesses will make more errors of commission (i.e., provide inaccurate information) and thus be less accurate as time passes are not well founded.

[2]This will not always be the case, as precision and quantity are often related, and a decrease in precision may result in a decrease in quantity. For example "He was wearing a jean jacket with a zipper" is both more precise and contains more information than "He was wearing a jacket".

The data show that when witnesses are able and willing to admit their lack of knowledge about an event, they are not likely to make such commission errors, even when their memories have weakened. Instead they choose to sacrifice informativeness to maintain accuracy. Due to this sacrifice, the errors that could be of concern after a delay are errors of omission (i.e., leaving out information).

Omission and commission errors may be more or less problematic depending on the goals of the questioner, or evaluator. If the main goal of an investigator is to obtain as much information as possible, omissions may be the more problematic error. If witnesses choose to leave out information in which they are not confident, the investigator will be left with fewer leads to follow. For example, a witness might not mention the suspicious car seen the morning of the crime because s/he is not sure if it was at the exact time of the crime or a little later. Similarly, if witnesses provide imprecise information because they do not feel they can provide specifics with confidence, this can also hold an investigation back. For example, a witness might describe the perpetrator as "tall" because s/he is not sure of an exact measurement. However, an estimate of 6'2'' may be more helpful to an investigator than "tall", even if the true height is 6'4''. In contrast, during a trial the goal of an attorney should be to provide the jury with accurate facts so that the correct verdict will be reached. Thus, errors of commission may be very undesirable from the perspective of an attorney. For example, if a witness testifies that the perpetrator was 6'2'' and used a black sedan to commit the crime, and the defendant drives a white sports car and is 6'4'' those mistake could hurt the witness's credibility and certainly do not help the jury make their decision. Thus, different error types may be of more or less concern depending on the context in which they are made.

Future studies should examine what beliefs jurors, attorneys, and investigators hold about the ways in which memory changes over time, and how they perceive witnesses who report that they "don't remember" many details. Common belief may be that witnesses who say "I don't know" often, or provide incomplete reports, have poor memories and therefore cannot be trusted. However, the data reported here indicate there is no relationship between saying "I don't know" and accuracy. Thus witnesses who say "I don't know", regardless of how strong their memories are, may not be any less accurate than other witnesses. They are simply monitoring their knowledge to ensure they do not provide incorrect information.

Effectiveness of metacognitive monitoring and control. The less correlated monitoring output is with actual accuracy, the less accurate reported information is (e.g., Kelley & Sahakeyn, 2003). This is because the control mechanism relies on the faulty information provided by the monitoring process. As witness confidence is the output of the metacognitive monitoring, anything that distorts confidence estimates may have interfered in some way with the monitoring process. In some cases this interference may occur during the initial monitoring process. For example divided attention at encoding decreases the effectiveness of the monitoring process (Kelley & Sahakeyn). Even after an initial assessment of confidence, variables that may be unrelated to actual accuracy could change confidence; for example, exposure to contradictory information.

One situation in which ineffective metacognition could affect witnesses' reports is after positive feedback has been given. Feedback has been shown to affect memory judgments in eyewitness identification studies (e.g., Wells & Bradfield, 1998). Similar feedback effects may be found in interviews, which would be problematic if witnesses

are interviewed repeatedly. If witnesses are given feedback that the information they provided was accurate, or useful, this may increase their confidence. Once confidence is inflated, during a subsequent interview the control process could allow more inaccurate information to be reported. Testifying in court would have essentially the same effect as another interview, and the same problematic overconfidence and higher rate of commission errors would be found in testimony provided.

Another potential influence on the effectiveness of the monitoring and control processes is the weakening of memory over time. Witnesses may become overconfident if metacognitive monitoring does not adjust adequately for a weakening memory trace after long periods of time. This overconfidence would result in control mechanism relying on inflated estimates of a memory's accuracy, and thus allowing inaccurate information to meet the accuracy threshold and be reported. The opposite is also possible: if witnesses over-correct for the effects of time on their memories then they could become underconfident, which would result in withholding accurate information. This may be a problem for the criminal justice system, as investigations can take months to complete, and trials often do not occur until long after an investigation is closed. Thus, future researchers should consider using a longer retention interval than was employed here to determine if these results hold.

Questioning Style. Ackerman & Goldsmith (in press) found that participants given a report option chose to provide a response only if they could meet both informativeness and accuracy criteria; otherwise they preferred to say "I don't know." Therefore, if witnesses do not mention certain aspects of an event in a free narrative or choose not to elaborate on them, this may be because they lack adequate knowledge to meet their

informativeness and accuracy criteria. Witnesses' ability to effectively employ metacognition to meet these criteria implies that questioners should be sensitive to witnesses' reasons for being imprecise or withholding information. In real-world interviews investigators often ask additional questions after witnesses have provided a free narrative, and these are likely to address details that were not mentioned, or aspects of the crime about which the witness provided little information. Following-up on unaddressed aspects of a crime with specific questions can be useful. For example, it provides retrieval cues which may elicit correct information that the witness simply forgot to mention. A problem with these questions however, is that they may effectively convert a free narrative interview to a cued or yes/no interview, which yield lower accuracy rates.

Follow-up questions may also have the unintended consequence of changing the witnesses' perceptions of their task. This may be beneficial if the questions convey to witnesses that a detailed report is desirable, which may not have been made clear at the outset of the interview. However, prodding witnesses to obtain more information if they initially say "I don't know" may imply that "I don't know' is not an acceptable response, despite what they previously believed or were told. Thus, follow-up questions may change the errors committed by witnesses from errors of omission to errors of commission. It would be valuable to know for future research what interviewees believe their task is, and examine how those perceptions vary depending on the way an interview is conducted. With the right instructions or questions, perhaps a balance can be struck in which more information is provided by the witness, but errors of commission are avoided because witnesses understand that they should provide as much information as possible,

but they should qualify what they say to reflect their uncertainty (cf Fisher & Geiselman, 1992).

The type of questions asked may also provide an implicit cue to witnesses indicating how precise responses should be. In the present experiment, participants in the cued condition provided more precise responses than participants in the free narrative condition. One possible explanation is that asking a relatively specific question (e.g., "How did the thief get away?") implies to interviewees that they should provide a precise response to this question ("He ran to a bike and rode away slowly"), whereas a broad question (e.g., "What happened?") implies that a broad answer is acceptable ("He got away on a bike"). The most basic example is when someone asks "How are you?"; the response to this broad questions is a similarly broad "Fine." In contrast, the more specific question "How was your trip?" will likely elicit a more elaborate answer. Thus implicit expectations may be sufficient to adjust the interviewee's informativeness and accuracy criteria (see Ackerman & Goldsmith, in press).

Two issues have been considered: asking follow-up questions, and the implicit expectation that respondents should provide precise answers to precise questions. Now consider the consequence of both principles acting together. Witnesses may have chosen to leave out information or be vague about details of which they are unsure. However, when asked follow-up questions they may come believe that they should not say "I don't know", and should provide a precise response. Thus, witnesses may choose to provide precise responses about details for which they do not have a strong memory. In such a situation, errors of commission are especially likely. Without the ability to adjust precision or withhold information, accuracy must decline.

Given the all way ways in which the question asked can influence the response provided, interviewers' recall of the questions they asked is very important (Hyman et al., 2007). Interviewers often write up reports which summarize the information provided, but do not specify the type of question that elicited each piece of information. Later in the investigation or during court proceedings such information may be valuable when evaluating witnesses' statements. For example, a witness might say during a free narrative "The thief was wearing a denim jacket". The same witness might not mention the thief's hair in that narrative, but in response to a specific follow-up question say "I think his hair was white." However, most likely the report will be written up as "The witness reported the thief had white hair and wore a denim jacket." Without knowing what question was asked, both pieces of information will be evaluated the same way, when in fact "denim jacket" is more likely to be correct than "white hair". Thus, it would be beneficial for reports to indicate whether information was provided during a free narrative or elicited in response to a more specific follow up question. This would likely require the recording of interviews, or more extensive note taking on the part of interviewers.

Implications for Researchers

The current findings also have implications for the design and evaluation of eyewitness memory research. Both the accuracy- and quantity-oriented approaches to memory research provide incomplete pictures. In order to understand memory, three important issues should be considered: accuracy, quantity and precision. For example, if a study were to find that a manipulation affected the accuracy rates of participants' reports, it would be important to consider whether there was a report option and whether

the precision of responses was under the participants' control. If not, then no conclusions about the trustworthiness of witness reports are truly justified as actual witnesses would generally be given control over report option and precision, and thus be able to improve their accuracy rates. Similarly if witnesses were found to provide a greater (or lesser) quantity of information in certain situations, this means little without considering the precision and accuracy of the information provided, and the conditions under which the information was elicited. To thoroughly evaluate witnesses' statements, and understand the effects of manipulations, researchers should take into account all three aspects of a response: quantity, accuracy, and precision.

Ideally, in the future a new measure of informativeness could be developed that would capture both quantity and precision of responses. For example, perhaps witness reports could be categorized as either "complete" or "incomplete" to capture the quantity of information provided, and "precise" or "vague" to capture precision. This would lead to four categories with different levels of overall informativeness. For example, "complete/precise" would be the highest level of informativeness, and incomplete/vague" would be the lowest. Another goal is the development of a system to evaluate the likely accuracy of reports based on the context of the interview (e.g., witness instructed that it's fine to say "I don't know", it was the first interview, or it was the fifth interview) and the types of question asked (e.g., free narrative or yes/no). This could be similar to the idea of Statement Validity Analysis (Stellar & Kohnken, 1989), with a checklist of interview characteristics to consider when assessing the reliability of a witness's report.

Limitations

Measuring precision, despite its complexity, was necessary to effectively explore precision adjustments in the context of an ecologically valid witness interview. Previously researchers examining precision only elicited information that could be easily and objectively scored as more or less precise, such as quantitative or color information. In contrast, in the current study participants were questioned about both quantitative and qualitative details, and were able to answer however they saw fit. Given the many different combinations of details and various phrasings in their responses, the precision of these responses could not be scored perfectly. Precision has the added drawback that it will almost always be confounded with the quantity of details provided. As more research is conducted researchers may define and measure precision in different ways. Should converging findings be reported despite the various imperfect measures of precision, this will provide strong evidence that these results are not a function of how precision was measured.

Another limitation of the current study is that no manipulation check was done to verify that the instruction manipulation was noticed or understood by the participants. Thus, we cannot know if null findings regarding the instructions were because the manipulation was ineffective or because people were unwilling or unable to change their criteria to provide more accurate, or inaccurate, information. I suspect that neither of the instructions deviated enough from the participants' initial expectations about their task to affect their behavior. However, instructions with the same message, but more emphatic wording, might be more effective (e.g. "Tell me everything you can remember, I don't care at all if it's right or not"). Previous research that manipulated incentives to be

accurate has shown that people can strategically control their responses to be more or less accurate. Thus is seems most likely that the relatively subtle instructions employed here did not provide sufficient motivation.

REFERENCES

Ackerman, R. & Goldsmith, M. (in press). Control over grain size in memory reporting – with and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology, 65*, 1119-1131.

Christiaanson, R.E. (1980). Prose memory: Forgetting rates for memory codes. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 611-619.

Conway, M. A., Cohen, G., & Stanhope, N. (1991). On the very long-term retention of knowledge acquired through formal education: Twelve years of cognitive psychology. *Journal of Experimental Psychology: General, 120*, 395-409.

Dorfman, J., & Madler, G. (1994). Implicit and explicit forgetting: When is gist remembered? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47a*, 651-672.

Ebbinghaus, H. (1895/1964). Memory: A contribution to experimental psychology. New York: Dover.

Ebbesen, E., & Rienick, C. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83*, 745-762.

Fisher, R. (1996). Implications of output bound measures for laboratory and field research in memory. *Behavioral and Brain Sciences, 19*, 197.

Fisher, R., & Geiselman, R. (1992). Memory-enhancing techniques for investigative interviewing: The cognitive interview. Springfield, IL: Thomas.

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

Goldsmith, M, Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language, 52*, 505-525.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 41-58). New York: Academic Press

Hyman, A., Schreiber, N., Abril, L., Pena, C., Carol, R., Rodriguez, E., & Diaz, V. (July, 2007). *Investigative interviewing and memory: How accurate are interviewers' recollections of investigative interviews after retention interval is manipulated?* Paper presented at the Seventh Biennial Meeting of the Society for Applied Research in Memory and Cognition, Bates College, Lewiston, ME.

Kelley, C., & Sahakeyn, L., (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48*, 704-721.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimmy, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*, 133-159.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology, General, 123*, 297-315.

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

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1095-1105.

Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology, 62*, 90-95.

McClelland, G. (2000). Nasty data: Unruly, ill-mannered observations can ruin your analysis. In H. Reis & C. Judd (Eds), *Handbook of research methods in social and personality psychology* (pp. 393-411) Cambridge: Cambridge University Press.

Pansky, A., & Koriat, A. (2004). The basic level convergence effect in memory distortions. *Psychological Science, 15*, 52-59.

Poole, D., & White, L. (1993). Two years later: Effects of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology, 29*, 844-853.

Stanhope, N., Cohen, G., & Conway, M.A. (1993). Very long-term retention of a novel. *Applied Cognitive Psychology, 7*, 239-256.

Steller, M., & Kohnken, G. (1989). Criteria-Based Content Analysis. In D.C. Raskin (ed.). *Psychological Methods in Criminal Investigation and Evidence* (pp. 217-245). New York: Springer-Verlag.

Weber, N. & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied, 14*, 50-60.

Wells, G.L., & Bradfield, A.L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*, 360-376.

Yaniv, I., & Foster, D. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology, General, 124*, 424-432.

Yaniv, I., & Foster, D. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making, 10*, 21-32.

APPENDIX A

So, keeping in mind the instructions you just received, please answer the following questions.

Did the thief look young?

Was the victim an older person?

Was the thief wearing khakis?

Was the thief wearing a blazer?

Did the thief take a purse?

Did the thief get away on a bike?

Did the thief have brown hair?

Was the thief acting unusual before the crime?

Was the woman carrying papers?

When the thief approached her, did the woman ignore him?

When the thief stole the bag, did he knock anything out of the victim's hands?

Were the thief's shoes brown?

Was the victim wearing brown?

Was the thief cleanshaven?

Was the victim's bag tan?

APPENDIX B

So, keeping in mind the instructions you just received, please answer the following questions.

Please describe, if you can, the thief's age.

Please describe, if you can, the victim's age.

Please describe, if you can, the pants was the thief wearing.

Please describe, if you can, the jacket was the thief wearing.

Please describe, if you can, what was stolen.

Please describe, if you can, how the thief got away.

Please describe, if you can, the thief's hair.

Please describe, if you can, how the thief was acting before the crime took place.

Please describe, if you can, what the victim was carrying, if anything.

Please describe, if you can, what the victim did when the thief approached her.

Please describe, if you can, how the thief stole the bag.

Please describe, if you can, the thief's shoes.

Please describe, if you can, the victim's clothing.

Please describe, if you can, the thief's face.

Please describe, if you can, the victim's bag.

APPENDIX C

So, keeping in mind the instructions you just received, please tell me everything that you remember about the crime that you witnessed. Please describe what you saw happen while also including descriptive details. For example, if you had seen someone steal a car, you should describe how the car was stolen AND what the car looked like. We have about 45 minutes left in this session, so go ahead and take as much time as you need.

APPENDIX D

INSTRUCTIONS – Accuracy / Yes-No

I'm about to ask you about the crime on the video you just watched. I want you to put yourself in the shoes of a witness to a crime, and to think of me as the detective who was assigned to investigate the case. You should assume that I know nothing about this crime, other than that someone stole something from someone. You should also assume that I am the only person working on this case and therefore there are limited resources to follow any leads that you provide- so it is very important that all the information you tell me is accurate, otherwise I might waste time following leads that end up being dead ends. Keep in mind that if you don't remember something you should not guess, instead just say "I don't know" or "I don't remember". I will be asking you questions that are yes/no questions, meaning, they can be answered by just saying yes, or just saying no. Please only say either yes or no, or "I don't know".

INSTRUCTIONS – Informativeness / Cued

I'm about to ask you about the crime on the video you just watched. I want you to put yourself in the shoes of a witness to a crime, and to think of me as the detective who was assigned to investigate the case. You should assume that I know nothing about this crime, other than that someone stole something from someone. You should also assume that you are the only witness to the crime, and so you are our only source of information. Therefore, I will need you to provide as much information as you possibly can, because you are all I have. Keep in mind that if you don't remember something you should not guess, instead just say "I don't know" or "I don't remember".

INSTRUCTIONS – Informativeness / Free Narrative

I'm about to ask you about the crime on the video you just watched. I want you to put yourself in the shoes of a witness to a crime, and to think of me as the detective who was assigned to investigate the case. You should assume that I know nothing about this crime, other than that someone stole something from someone. You should also assume that you are the only witness to the crime, and so you are our only source of information. Therefore, I will need you to provide as much information as you possibly can, because you are all I have. Keep in mind that if you don't remember something you should not guess, instead just leave out the information.

APPENDIX E

	QUANTITY	ACCURATE	inaccurate	unscorable	Could	PRECISION	uncertainty	age	IDK	discuss	comments
Q1											
Q2											
Q3											
Q4											
Q5											
Q6											
Q7											
Q8											
Q9											
Q10											
Q11											
Q12											
Q13											
Q14											
Q15											
other 1											
other 2											
other 3											
other 4											

Scorer _____

PPT # _____

Condition _____

VITA

JACQUELINE EVANS

September 22, 1982	Born, Sierra Vista, Arizona
2004	B.A., Cognitive Science University of Virginia Charlottesville, Virginia
2007	M.S., Psychology Florida International University Miami, Florida
2004-2007	Graduate Assistant Florida International University Miami, Florida
2007-2008	Instructor Florida International University Miami, Florida

PUBLICATIONS AND PRESENTATIONS

Evans, J.R., Danielsen, E., & O'Neil, K.M. (2005, March). *The role of deadlock instructions and parole alternatives in capital trial verdicts*. Paper presented at the American Psychology-Law Society 2005 Conference, La Jolla, CA.

Evans, J. R., Marcon, J. L., & Meissner, C. A. (in press). Cross-racial lineup identification: The potential benefits of context reinstatement. Psychology, Crime, & Law.

Evans, J. & O'Neil, K. (March, 2007). *Capital juror sentencing decisions: Do sentencing options, deadlock consequences, and comprehension matter?* Paper presented at Off the Witness Stand: Using Psychology in the Practice of Justice, John Jay College of Criminal Justice, New York City.

Evans, J.R., & O'Neil, K. (March, 2006). *Jurors' evaluations of mitigating evidence of mental illness or impairment.*. Paper presented at the American Psychology-Law Society Conference, St. Petersburg, FL.

Evans, J.R., O'Neil, K.M, & Mazzone, J. (March, 2006). *Downloading and sharing music via the Internet: An empirical study of the factors influencinging downloading behavior*. Poster presented at the American Psychology-Law Society Conference, St. Petersburg, FL.

Evans, J.R. & Schreiber Compo, N. (revisions submittted). Mock jurors' percpetions of identifications made by intoxicated eyewitnesses.

Evans, J. & Schreiber, N. (March, 2007). *Intoxicated witness and victims of crime: How common are they?* Poster presented at Off the Witness Stand: Using Psychology in the Practice of Justice, John Jay College of Criminal Justice, New York City.

Evans, J.R., Schreiber Compo, N., & O'Neil, K.M. (March, 2008). *Mock Jurors' Evaluations of Victim Fault across Different Crimes and Intoxication Levels*. Poster presented at the American Psychology-Law Society Conference, Jacksonville, FL.

Hernandez, E., Schreiber Compo, N., Evans, J. R., Carol, R., Diaz, V., Rodriguez, E., Ortega, L., Sylvester, K., Charosky, M., Kemp, D., Lobos, A., Stauffer, D., Wood, S., Perry, A., Villalba, D., & Rose, S. (March, 2008). *Eyewitness memory and intoxicated witnesses: Are they more confident than sober or placebo witnesses?* Paper presented at the American Psychology-Law Society Conference, Jacksonville, FL.

O'Neil, K.M., Danielsen, E., Calia, T., Reardon, M., Lawson, K., York, R., Evans, J.R., Perez, V., & Penrod, S. (March, 2006). *Jurors and products liability: How deliberations change jurors' focus*. Paper presented at the American Psychology-Law Society Conference, St. Petersburg, FL.

Perez, V., Evans, J.R., & O'Neil, K.M. (March, 2006). *Future dangerousness: The impact of risk factors and the GBMI option in insanity cases*. Paper presented at the American Psychology-Law Society Conference, St. Petersburg, FL.

Schreiber, N., Evans, J., Ham, L., Garcia, T., Rose, S., Pignatello, V., Carol, R., Perry, A., Villalba, D., Diaz, V., Rodriguez, E., Turner, G., Abril, L. & Alonso, T. (July, 2007). *Intoxicated witnesses: Are they more vulnerable to misinformation?* Paper presented at the Seventh Biennial Meeting of the Society for Applied Research in Memory and Cognition, Bates College, Lewiston, ME.

Wright, D., Carlucci, M., Evans, J. & Schreiber, N. (July, 2007). *The impact of eyewitness identifications from simultaneous/sequential and blind/non-blind lineups*. Paper presented at the Seventh Biennial Meeting of the Society for Applied Research in Memory and Cognition, Bates College, Lewiston, ME.

York, R., Evans, J.R., & O'Neil, K.M. (2006, March). *The CSI effect: Presentation style, evidence quality, and a possible remedy*. Paper presented at the American Psychology-Law Society Conference, St. Petersburg, FL.

York, R., Evans, J., & O'Neil, K. (2005, March). *Gender effects and the mediating mechanism of self-referencing in the trials of battered spouses*. Paper presented at the American Psychology Law Society 2005 Conference, La Jolla, CA