8-2018

Accuracy differences between interview formats : examining factors that may influence metamemory utilization.

Robin F. Hopkins
University of Louisville

ACCURACY DIFFERENCES BETWEEN INTERVIEW FORMATS: EXAMINING

FACTORS THAT MAY INFLUENCE METAMEMORY UTILIZATION

By

Robin F. Hopkins
B.S., High Point University, 2012
M.S., University of Louisville, 2015

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Experimental Psychology

Department of Psychological and Brain Sciences
University of Louisville
Louisville, Kentucky

August 2018

ACCURACY DIFFERENCES BETWEEN INTERVIEW FORMATS: EXAMINING

FACTORS THAT MAY INFLUENCE METAMEMORY UTILIZATION

By

Robin F. Hopkins
B.S., High Point University, 2012
M.S., University of Louisville, 2015

A Dissertation Approved on

July 30, 2018

by the following Dissertation Committee:

_____
Dissertation Director
Keith B. Lyle

_____
Marci S. DeCaro

_____
Judith H. Danovitch

_____
John R. Pani

_____
Thomas Hughes

ABSTRACT

ACCURACY DIFFERENCES BETWEEN INTERVIEW FORMATS: EXAMINING

FACTORS THAT MAY INFLUENCE METAMEMORY UTILIZATION

Robin F. Hopkins

July 30, 2018

Eyewitnesses are important for criminal investigations. When interviewing

witnesses, police first ask witnesses to describe what they observed. This initial

statement is a free narrative (FN) and is usually highly accurate. Next, police ask follow-

up questions (FQs), which generate additional information but often at the cost of

accuracy. This dissertation aims to examine factors that may contribute to the effect of

interview format on accuracy and whether FQ instructions change detail quality provided

in follow-up question responses (FQRs). Additionally, individual differences in

responding to FQ instructions were explored.

Subjects watched a movie clip depicting a crime. Next, subjects filled out

personality questionnaires and then typed their FNs. Subjects were then given one of five

FQ instructions that encouraged accuracy (Accuracy+), encouraged informativeness

(Informative+), encouraged "I don't know" responses if needed (IDK+), discouraged "I

don't know" responses unless absolutely necessary (IDK-), or only told them they would

be asked FQs (Control). Ten variables were used to measure interview quality (e.g.,

accuracy, precision, quantity).

FN quality was compared to FQR quality in the Control condition to determine

the effect of interview format, without influence of instructions. FQR quality was

compared across FQ conditions to determine the effect of FQ instructions. FQR quality

was correlated with conscientiousness and social desirability scores to explore individual

differences. Analyses were conducted once using all FQRs and once using only

responses if the corresponding FQ topic had been mentioned during the preceding FN.

Interview format affected all variables except overall accuracy. This null effect

on accuracy does not align with previous interview literature and limited the

interpretation of other effects. IDK+ instructions were the only instructions that affected

FQR quality. IDK+ subjects said "I don't know" significantly more frequently, and

provided significantly fewer inaccurate details, in FQRs than Control subjects. Yet,

IDK+ subjects were not significantly more accurate than Control subjects. Thus, it

remains unclear whether IDK+ instructions selectively filter inaccurate details or reduce

total detail output. The effects of IDK+ instructions were only significant when using all

FQRs. Significant individual differences emerged in Accuracy+, Informative+, and

Control conditions when using all FQRs.

TABLE OF CONTENTS

PAGE

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Overview

Ensuring the quality of eyewitness statements is important for identifying, finding, and prosecuting suspects. Police are usually the first people to take thorough statements from eyewitnesses, when memories are still fresh and minimally contaminated. When there is little physical evidence in a case, eyewitness accounts may be the best, or only, available evidence for correctly identifying and prosecuting criminals. Especially in cases where eyewitnesses are the primary source of evidence, the information they provide forms the basis of decisions which drive the course of the investigation. Thus, the details eyewitnesses provide can lead to serious consequences for suspects who are identified and located. To be of maximal value, eyewitness accounts should contain details that are numerous, specific, and (most importantly) accurate. Therefore, the challenge for police is to interview witnesses in a manner that elicits as many details as possible, and details that are precise, while simultaneously ensuring that those details are accurate.

Police typically use a standard format during their first interview with witnesses (Inbau, Reid, Buckley, & Jayne, 2015; Yullie & Cutshall, 1986). First, witnesses are simply asked what happened. The initial statement can be considered a free narrative

1

(FN) because there are no specific instructions regarding what to recall, other than that it should be about the event witnessed. The content of FNs is usually highly accurate, but provides an incomplete account of what happened (Hilgard & Loftus, 1979; Inbau et al., 2015). Therefore, the standard interview format continues with follow-up questions (FQs) to fill in gaps.

FQs help law enforcement gather missing information, but, critically, the accuracy of the details provided in FQ responses (FQRs) is consistently significantly lower than the details provided during FNs (Evans & Fisher, 2011; Hilgard & Loftus, 1979; Lipton, 1979; Odinot, Wolters, & van Koppen, 2009; Taylor & Dando, 2018; Yarmey & Yarmey, 1997). Police would benefit from understanding the specific psychological phenomena that result in a higher proportion of inaccurate details during FQRs in order to improve interview techniques and know how to best use the information gathered during interviews to find and successfully prosecute the correct person.

Purpose of Dissertation

This dissertation aims to examine factors that may contribute to the effect of interview format on accuracy. First, I will compare the quality of details provided in FNs to the quality of details provided in FQRs to determine what types of errors are contributing to differences in accuracy. In order to isolate the effect of interview format, no influential instructions will be used in this initial comparison. Additionally, I will examine the effect of potentially influential FQ instructions on the quality of FQRs. Finally, I will explore the relationship between individual differences and quality of FQRs, as people may respond differently to potentially influential FQ instructions.

Literature Review

2

Definition of Metamemory

When people answer questions based on their memory, a few steps occur between the retrieval of details and a response being provided (Ackerman, & Goldsmith, 2008; Goldsmith, Koriat, & Pansky, 2005; Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Koriat & Goldsmith, 1996). This decision-making process is illustrated in Figure 1. First, a potential detail must be retrieved from memory (e.g., a car was black). Next, this detail is evaluated based on two separate criterions. Respondents judge how confident they are about the accuracy of each detail and the informativeness of the detail, which is how useful they perceive the information to be (Ackerman & Goldsmith, 2008). For example, someone could be completely confident in the accuracy of describing the color of a car as "within the spectrum of visible light," but the respondent would not bother to report this description because it is so vague as to be useless. On the other hand, someone may be only moderately confident in the accuracy of describing the color of a car as a specific color, but because the respondent feels that piece of information would be helpful, it is included in the description. The criterion levels for accuracy and informativeness are set by the respondent, but can be influenced by perceived or explicit expectations of the questioner, such as providing monetary incentives for accurate responses (Ackerman & Goldsmith, 2008; Goldsmith, Koriat, & Pansky, 2002; Koriat & Goldsmith, 1994; Scoboria & Fisico, 2013). After people have judged the perceived accuracy and informativeness of a detail, the detail is either provided or withheld. If the initial detail is rejected because a respondent is not sufficiently confident about its accuracy, or judges it to be insufficiently informative, the entire process may be repeated with a different potential detail that is either more general (e.g., the car was a dark color)

or more specific (e.g., the car was black matte). This process of self-monitoring is called

metamemory. How people use metamemory determines which details are reported, and

at what level of precision.
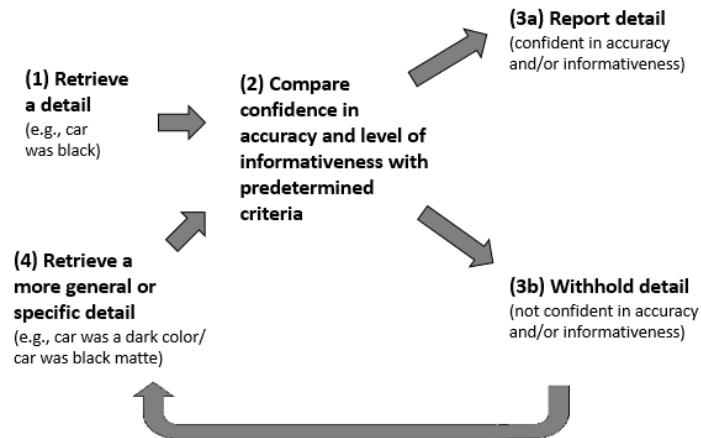
Figure 1. Illustration and example of metamemory steps. Availability of report option is illustrated by steps 3a and 3b. Control over grain size is illustrated by steps 1 and 4.

Three aspects of metamemory are of interest in this dissertation: availability of

report option, control over grain size, and how perceived and explicit expectations of the

questioner affect how people utilize control over report option and grain size.

Availability of report option simply means a respondent has the option to report or

withhold a detail. This is illustrated by steps 3a and 3b in Figure 1. Responding under

forced recall conditions means that the option to withhold details is removed, whereas

responding under free recall conditions means that the option to withhold details is

available. Control over grain size means that a respondent has the option to respond

using a precise description (i.e., fine-grain detail) or a general description (i.e., coarse-

grain detail). This is illustrated by steps 1 and 4 in Figure 1.

Definition of Quality of Details

There are three distinct ways in which detail quality is commonly measured: accuracy, quantity, and precision. Accuracy is the proportion of correct details out of the total number of details. Accurate details are necessary for investigative purposes, because inaccurate details may cause the police to investigate incorrect leads. Following incorrect leads can waste time, and possibly lead to a false conviction. Quantity of details is the sum of discrete details provided (both accurate and inaccurate). A high quantity of accurate details is ideal for investigative purposes, as additional details narrow down the list of suspects. Precision is the level of specificity at which a detail is reported. Precise, or fine-grain, accurate details are ideal for investigative purposes; similar to quantity of details, more precise details narrow down the list of suspects.

Quantity and precision are distinct measurements of details, as the same object can be described with varying levels of precision without adding to the quantity of details, and vice versa. For example, a vehicle can be described as a car, sedan, Toyota, or Corolla. Each of these descriptions varies in the level of precision, but each description would only contribute one discrete detail to the quantity of details. However, both measurements contribute to the overall informativeness of details provided. A high quantity of details, or specific details, or a combination of both, can all contribute to an informative statement to police, which is ideal for successfully investigating a crime.

Review of Metamemory Literature

A two-phase design is typically used to study the effect of freedom to utilize metamemory (i.e., availability of report option and control over grain size) on the quality (i.e., accuracy, quantity, precision) of details produced. Subjects are asked the same questions in each phase. When this design is used to study the effect of availability of

5

report option, subjects are required to provide answers for each question (even if it is just a guess) during the first phase, and they have the freedom to choose which questions to answer during the second phase in order to increase their accuracy. For example, if asked when President John F. Kennedy was assassinated, subjects would have to provide an answer during the first phase, but they could choose whether or not to provide an answer during the second phase. When this design is used to study the effect of control over grain size, subjects are required to provide both a fine-grain and a coarse-grain answer during the first phase, and they have the freedom to choose which grain size to provide during the second phase in order to increase accuracy. For example, if asked when Kennedy was assassinated, after providing both the exact year (i.e., fine grain) and a ten-year range (i.e., coarse grain) during the first phase, subjects would have the freedom to choose between providing the fine-grain or coarse-grain answer during the second phase.

Studies examining the role of availability of report option have found an accuracy-quantity tradeoff between the two phases (Koriat & Goldsmith 1994, 1996; Koriat, Goldsmith, Schneider, & Nakash-Dura, 2001), which is illustrated in Figure 2A. During initial forced recall phase, because every question must be answered, a high quantity of details is provided. However, because subjects have to answer questions even if all they can provide is a guess, accuracy is relatively low during this phase. During the following free recall phase, fewer questions are answered, but the accuracy of the answers provided is relatively high.

Studies examining control over grain size have found an accuracy-precision tradeoff between the two phases (Ackerman & Goldsmith, 2008; Goldsmith, Koriat, & Pansky, 2005; Koriat, Goldsmith, & Weinberg-Eliezer, 2002), which is illustrated in

6

Figure 2B. In these studies, the quality of details (i.e., accuracy and precision) of the

fine-grain details provided during the first phase is compared to the quality of details of

the *chosen* grain size from the second phase. Precision level is (by necessity) high in the

fine-grain details of the first phase, but overall accuracy is low. Precision level is lower

in the second phase when people have control over grain size, however overall accuracy

is relatively high.
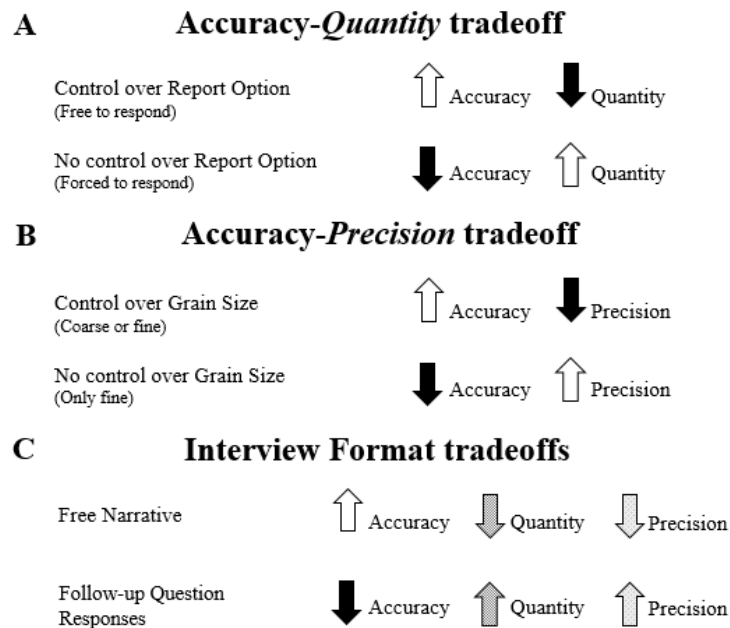
Figure 2A-C. Illustrations of the similarities between the accuracy-quantity (A) and
accuracy-precision (B) tradeoffs found in the metamemory literature and the same
tradeoffs between interview formats (C) in the interview procedure literature. The
changes in quantity and precision between interview formats are colored gray to represent
the current strength of evidence of these patterns.

Review of Interview Literature

Accuracy is typically high ($Ms \geq 90\%$) during FNs, and significantly lower during

FQRs ($60\% \leq Ms \leq 90\%$) (e.g., Dietze, Sharman, Powell, & Thomson, 2013; Evans &

Fisher, 2011; Lipton, 1977; Odinot et al., 2009; Taylor & Dando, 2018). However, the

quantity of details provided in FQRs is usually higher than in FNs (e.g., Dietze et al., 2013; Lipton, 1977; Sharps, Herrera, Dunn, Alcala, 2012; Taylor & Dando, 2018; Yarmey & Yarmey, 1997). In other words, there is typically an accuracy-quantity tradeoff between the two interview formats. Two studies report the number of details provided in each interview format, but did not test whether there was a statistical difference in quantity (Evans & Fisher, 2011; Odinot et al., 2009). Subjects in both of these studies appear to have produced an equal number of details in both interview formats, yet the accuracy of their FNs was significantly higher than their FQRs. Thus, while the majority of the evidence suggests there is an accuracy-quantity tradeoff between interview formats, this finding may not be consistent in the literature. Only one study has examined whether precision differs as a function of format and it found that details were marginally more precise in FQRs than in FNs (Evans & Fisher, 2011). Thus, while there is fairly strong evidence for an accuracy-quantity tradeoff, there is only weak evidence of an accuracy-precision tradeoff. The two tradeoff types are depicted graphically in Figure 2C.

Instructions that Influence Accuracy

Usage of metamemory can be influenced by incentives to maintain high accuracy, suggesting that people can, and will, utilize metamemory differently depending on the situation. Studies in the metamemory literature show that subjects who are incentivized to provide accurate answers require a higher level of confidence in the accuracy of a detail before they decide to report it (during the second phase) than when there is no accuracy incentive (e.g., Koriat & Goldsmith, 1994, 1996). This higher accuracy criterion results in fewer details being reported, but accuracy generally increases.

Apparently people are taking advantage of their report option to withhold details they would otherwise report in the absence of an accuracy incentive.

Explicitly informing people that they have the option to say "I don't know" in response to questions also has an effect on accuracy (Ackerman & Goldsmith, 2008; Scoboria & Fisico, 2013, Experiment 2). Scoboria and Fisico instructed subjects in one of three different ways, and then asked a series of questions about a witnessed event. The control group did not receive any explicit instructions regarding whether "I don't know" was an acceptable response. One experimental group was encouraged to respond "I don't know" if needed, and another experimental group was discouraged from responding "I don't know" unless absolutely necessary. The group that was encouraged to use "I don't know" provided details that were significantly more accurate than in the control group and the group discouraged from using "I don't know". One interpretation of this finding is that, unless people are explicitly told that "I don't know" is acceptable, people assume that they should provide answers, even if unsure—in other words, they should minimize use of report option.

After answering questions, subjects in Scoboria and Fisico (2013) clarified what they meant by each "I don't know" statement. Almost all of the clarifications fell into one of three categories: 1) subjects did not remember anything about that detail; 2) subjects remembered something, but were not confident in their answer; or 3) subjects remembered something, but it was so general they assumed it was not useful (i.e., uninformative). These reasons make it clear that responding with "I don't know" reflects utilization of metamemory. When people chose to respond "I don't know" because they did not remember anything about that detail or because they were not confident in their

answer, they were taking advantage of availability of report option. When people chose to respond "I don't know" because they remembered something but perceived it to be uninformative, they utilized both report option and control over grain size. It therefore appears that Scoboria and Fisico's instructions influenced utilization of metamemory, which in turn affected accuracy.

Applying Metamemory Findings to Quality of Interviews

Tradeoffs observed in the metamemory literature resemble those observed in the literature on interview procedures, but few researchers have used open-ended questions (as opposed to close-ended questions that have one correct answer) based on personal experiences (as opposed to general knowledge) to examine metamemory differences. Yet, just like with close-ended questions used in the metamemory literature, quantity and precision of details provided during interviews can be used to understand how report option and control over grain size are being utilized, respectively. Specifically, if more details are provided in response to one interview format compared to the other, it would suggest that interview format influences how people utilize their report option. Likewise, if more precise details are provided in response to one interview format compared to the other, it would suggest that interview format influences how people utilize their control over grain size.

Evans and Fisher (2011) used open-ended questions based on a personal experience to study metamemory utilization during interviews (including FN and FQ formats), however their primary goal was to examine how metamemory is utilized in immediate and delayed interviews. They found that, compared to an immediate interview, people provide fewer details and less precise details in a delayed interview.

10

While the accuracy of immediate interviews was marginally higher than the accuracy of delayed interviews, subjects were able to maintain a relatively high accuracy rate at both time points (Ms > .89). Subjects reported few and imprecise details in the delayed interview, which resulted in a relatively high accuracy rate, suggesting they were utilizing their report option and control over grain size, respectively, based on an accuracy criterion. Additionally, after subjects had finished answering the FQs, Evans and Fisher requested that subjects provide an answer for the FQ for which they had initially given "I don't know" responses. In other words, subjects initially answered FQs with control over report option, and then had to answer them without a report option. The accuracy of FQRs provided when subjects had control over report option was significantly higher than when subjects had no report option, suggesting that they were making report option decisions to increase accuracy. These results, too, indicate that people utilize both availability of report option and control over grain size to increase accuracy during interviews. Notably, this research also provides evidence that metamemory utilization can be studied using open-ended questions and using personal experiences.

While Evans and Fisher (2011) demonstrated that people utilize metamemory during interviews to achieve high accuracy, their study did not thoroughly examine metamemory utilization differences caused by interview format. However, they do report format differences in accuracy and marginal differences in precision level, in a pattern that demonstrated an accuracy-precision tradeoff with accuracy being highest during FNs. A more comprehensive study focused on the effect of interview format on accuracy, rather than the effect of interview timing, is necessary. Regardless of when an

interview is conducted (immediately or delayed), it is likely that FQs will be asked, and low accuracy associated with this interview format needs to be explored.

In addition to focusing on interview format effects, the coding guidelines used in Evans and Fisher (2011) could be improved upon to gain a more in-depth understanding of how grain-size decisions are made. Rather than only coding for the average precision level of details in each interview format like Evans and Fisher, it would reveal more about people's decision-making process to code for the precision level of accurate and inaccurate details within each format. Evans and Fisher found precision in both interview formats to be, on average, moderate, but this result could stem from multiple different scenarios. For example, if accurate details provided in one interview format were extremely general and inaccurate details were extremely specific, this would result in a moderate average precision level. Alternatively, if the accurate and inaccurate details provided in one interview format were all moderately informative (neither general nor specific), this would also result in a moderate average precision level. Because Evans and Fisher only reported average precision level, collapsed across detail accuracy, we cannot know exactly what was going on in each format. Coding for precision of accurate and inaccurate details separately would clarify whether an effect of interview format on precision level is primarily driven by specific inaccurate details or specific details that are both accurate and inaccurate.

Additionally, there is evidence in the metamemory literature that instructions can influence how metamemory is utilized, yet FQ instructions used in interview procedure studies have been highly variable. Some studies encouraged accuracy (e.g., Lipton, 1977), some encouraged both accuracy and informativeness (e.g., Cady, 1924; Marquis,

Marshall, & Oskamp; 1972), and some include explicit instructions regarding "I don't know" responses (e.g., Evans & Fisher, 2011). The consistency of the effect of interview format on accuracy suggests that accuracy differences can be attributed to interview format, not FQ instructions. However, the possible inconsistency of the effect of interview format on quantity suggest that FQ instructions could influence how people are utilizing their report option when responding to FQs. Due to the minimal research that measures precision of details, it is also unclear what effect different FQ instructions may have on how people utilize their control over grain size.

Further, assuming FQ instructions do significantly alter how people utilize metamemory when providing FQRs, it would be helpful to know whether aspects of people's personalities lead them to make more extreme criterion changes, which could lead to either more or fewer errors during FQRs. Social desirability and conscientiousness are personality characteristics that could potentially influence how strictly a person follows instructions. People who score high on the social desirability scale are likely to respond in ways that will make others view them favorably (Crowne & Marlowe, 1960). Conscientiousness is associated with behaviors that include deliberation, dutifulness, and self-discipline (John & Srivastava, 1999), and has been linked to being rigidly attached to goals (Nettle, 2006).

If people believe that either accuracy or informativeness is most important in deciding what to report in their FQRs, and they want to appear helpful or are rigidly attached to this goal, it could influence how they choose to utilize their metamemory. For example, when instructions emphasize accuracy, people who are highly conscientious or wish to obtain approval may be especially likely to make use of report option, at the

expensive of informativeness. Similarly, when instructions emphasize informativeness, these same individuals may be more likely to use metamemory in ways that maximize informativeness, at the expense of accuracy.

In addition to the inconsistency of FQ instructions, previous studies differ in how FQ questions are chosen, which creates another issue for comparing results across the interview literature. Some studies use pre-determined sets of questions to ask everyone, regardless of what topics are discussed in a person's FN (e.g., Evans & Fisher, 2011; Lipton, 1977), while other studies have trained interviewers deciding which FQs to ask each person based on what they report in their FN (e.g., Memon, Wark, Bull, & Koehnken, 1997; Odinot et al., 2009; Taylor & Dando, 2018). Police are unlikely to ask an FQ on a topic that has not yet been brought up in a person's FN. For example, police are unlikely to ask "can you tell me more about the gun?" if a witness has not mentioned that the perpetrator had a gun. Therefore, when a pre-determined set of questions is used in an experiment, two set of analyses should be conducted in order to determine whether results are robust, regardless of whether FQ topics were mentioned in a person's FN. One set of analyses should include all FQRs, and the other set of analyses should include on FQRs if the corresponding FQ topic was mentioned in the preceding FN.

Implications

A small percentage of crimes result in arrests (Center for the Study and Prevention of Violence, 1995), and of those arrests, a small percentage end with a conviction (Blumstein & Cohen, 1979). Police investigators' ability to solve crimes greatly depends on the information available to them (Skogann & Antunes, 1979). A better understanding of how people utilize metamemory when providing FNs and FQRs

could lead to improvements in investigative procedures. Specifically, investigators could adjust how they follow-up on information provided by eyewitness. For example, suppose an officer is told that a perpetrator drove away from the scene during a witness's FN, and then asks the witness whether he or she remembers any information about the car, and is told that the car was black. If that officer knows that people are more likely to report inaccurate and precise details during FQRs, then the officer may decide to alert the public to look for a dark-colored car, rather than the more specific (but potentially incorrect) description of a black car. In contrast, if an officer is told that a perpetrator drove away from the scene in a black car during a witness's FN, and knows that people are more likely to report accurate details in FNs—regardless of precision size—then the officer may decide that it is acceptable to tell the public the specific description of the car.

Furthermore, understanding how instructions influence utilization of metamemory when responding to FQs could help the police tailor the wording of instructions used before asking FQs. The wording of instructions would likely depend on the nature of the crime. If there is other evidence to corroborate what an eyewitness says (e.g., multiple witnesses, physical evidence, nearby surveillance camera), then it may be okay to encourage informative details and it may be unnecessary to explicitly mention that "I don't know" is an acceptable response. However, there are some crimes where eyewitness evidence is the only, or the most valuable, evidence in a case. For example, when interviewing someone who has been a victim of abuse that is not reported immediately after the event, it is extremely important that the highest level of accuracy is maintained because there would likely no corroborating evidence available, even if it comes at the cost of informativeness. For these types of crimes, it could be extremely

important to explicitly tell the eyewitness that "I don't know" is an acceptable response, or to encourage accurate responses.

Current Experiment

Subjects were assigned to one of five conditions. In each condition, subjects first watched a movie clip, then produced an FN, and finally answered FQs. Conditions differed only in the instructions preceding the FQs. In the control condition subjects were simply told that they would be asked questions about the event witnessed in the movie clip. Instructions in the four experimental conditions included additional information (see Appendix A for all instructions). In one condition subjects were encouraged to respond "I don't know" when needed (IDK+) while in another condition subjects were discouraged from responding "I don't know" unless absolutely necessary (IDK-). In another condition subjects were encouraged to be as accurate as possible (Accuracy+) while in another condition they were encouraged to be as informative as possible (Informative+).

Details produced by respondents were coded for accuracy, quantity, and precision. Accuracy was determined by the number of accurate details divided by the total number of details (accurate + inaccurate). Quantity of details was defined as the total number of discrete units of information recalled. Precision of details was coded on a 1 (*very general*) to 5 (*very specific*) scale. For example, "Stephen" was coded as one discrete detail, while "one of the brothers" was coded as two discrete details. In regards to precision level, "Stephen" was coded as a level 5, and "one of the brothers" was coded as a level 4 because it is still more specific than saying a "boy"(3), "child" (2), or "person" (1).

Figure 3 illustrates how details in each interview format were measured. Accuracy, total quantity of details, and overall precision level were analyzed to test for two possible tradeoffs between interview formats: accuracy-quantity or accuracy-precision. An accuracy-quantity tradeoff would mean that subjects provided few details that were highly accurate in one interview format, and provided many details that were not highly accurate in the other interview format. An accuracy-precision tradeoff would create the same pattern but based on accuracy and precision rather than accuracy and quantity. Exhibiting both tradeoffs simultaneously would mean that subjects provided both a larger quantity of details and more precise details at the expense of being less accurate; I call this an accuracy-informative tradeoff.

Quantity of accurate details, quantity of inaccurate details, and quantity of incorrect-category inaccurate details were used to gain insight into whether utilization of report option differs depending on interview format or FQ instructions. Incorrect-category inaccurate details are details that do not fall in the same general category as the correct description. An example of this type of detail would be describing a car as blue when it was actually white; white and blue do not belong to the same general category of "dark color" or "light color."

Precision level of accurate details, precision level of inaccurate details, and quantity of correct-category inaccurate details were used to gain insight into whether utilization of grain size differs depending on interview format or FQ instructions. Correct-category inaccurate details are details that fall in the same general category as the correct description. An example of this type of detail would be describing a car as blue

when it was actually black; both blue and black belong to the same general color category

of "dark."
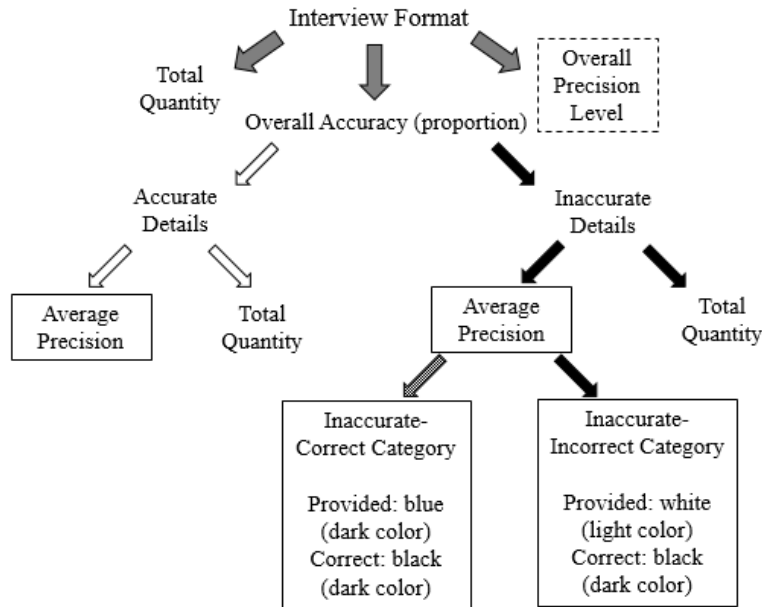
Figure 3. Illustration of how details in each interview format were measured. Measurements with boxes around them indicate variables which have not been used in previous studies, or in the case of overall precision level, in only one previous study.

In addition to the dependent variables coded for in each interview format, the

number of "I don't know" responses provided in FQRs was used as an additional measure

of people using their report option—by explicitly withholding information—depending

on instructions. Very few, if any, instances of saying "I don't know" were expected

during FNs, thus this response was only coded for in FQRs.

Social desirability and conscientiousness were measured using questionnaires to

examine individual differences in the effect of FQ instructions on utilization of

metamemory.

The quality of details produced in FQRs were analyzed in two ways. First, all

FQR details from all six FQs were included—regardless of whether the FQ topic was

brought up in each individual's FN (FQ All). Next, the FQR details were included based on FQ topics mentioned in each individual's FN (FQ Mention). For example, different FQs in the study request more information about a young girl and a drycleaner. Suppose a subject mentions the girl in his or her FN, but does not mention the drycleaner. One set of analyses use all FQRs (details about the drycleaner would be included), whereas the second set of analyses only used details from FQ topics mentioned in each person's FN (only details about the girl would be included).

Research Questions and Hypotheses

This dissertation aims to answer two key questions: 1) When no metamemory instructions are provided prior to FQs, do the details reported in FNs and FQRs differ in quality (i.e., accuracy, quantity, and precision)? 2) Does the provision of explicit metamemory instructions prior to FQs influence the quality of details reported in subsequent FQRs?

While I will use ten dependent variables to measure the quality of details, four of them (correct category-inaccurate details, incorrect category-inaccurate details, precision of accurate details, and precision of inaccurate details) have not been examined in the previous studies. The following seven hypotheses are limited to the variables previously examined in research.

The majority of research in both the metamemory and interview literatures indicates that the number of accurate details is relatively independent of interview format and availability of report option, with differences in the proportion of correct information being driven by the number of inaccurate details. Therefore, I anticipate an effect of both interview format and FQ instructions on inaccurate details without a corresponding effect

on accurate details. People typically provide very few inaccurate details in FNs, so I expect that subjects will provide significantly fewer inaccurate details in FNs than in Control condition FQRs. When subjects are asked FQs, they may interpret these questions as a request for more information than they provided during their FNs. This interpretation should be reinforced by IDK- and Informative+ instructions, but contradicted by Accuracy+ and IDK+ instructions. I therefore expect that subjects in the Control, IDK-, and Informative+ conditions will report significantly more inaccurate details in their FQRs than subjects in the Accuracy+ and IDK+ conditions. I also anticipate that the number of "I don't know" responses will differ depending on metamemory instructions, with subjects in the IDK+ condition responding "I don't know" more frequently than subjects in the IDK-, Informative+, and Control conditions. Accuracy+ instructions may also encourage subjects to say "I don't know," but subjects can also increase accuracy by reporting coarse-grained information (as opposed to fine-grain), therefore no hypotheses concerning "I don't know" frequency in the Accuracy+ condition will be made.

The total quantity and overall accuracy of details are both calculated using a combination of the inaccurate and accurate details provided during an interview. Therefore, my predictions for total quantity and overall accuracy mirror those for the quantities of accurate and inaccurate details. I expect that subjects' FNs will be significantly more accurate and less detailed than their FQRs. FQRs of subjects in the IDK-, Informative+, and Control conditions will be significantly less accurate and more detailed than subjects' responses in the Accuracy+ and IDK+ conditions. In terms of precision, subjects may feel a need to be more precise when asked FQs, versus when

asked to provide an FN. Therefore, FQRs should be more precise than FNs in the Control condition. Instructions that reinforce subjects' perception of an implicit request for more informative details (Informative+) should also lead to more precise details. The Accuracy+ instructions do not explicitly refer to precision, but one method by which people can increase accuracy is by providing less precise details. Following this reasoning, I anticipate that subjects in the Accuracy+ conditions will report significantly less precise details than those in the Informative+ and Control conditions. In other words, I anticipate that the Accuracy+, IDK+, and IDK- instructions will exhibit an accuracy-quantity tradeoff between interview formats. I expect the Informative+ and Control conditions to demonstrate an accuracy-informative tradeoff.

A third question posed in this dissertation is exploratory: Do subjects who score high on the Social Desirability Scale and/or the conscientiousness subset of the Big Five Inventory provide FQRs of a different quality than subjects who have low scores on these tests? People who score high on social desirability or conscientiousness may be influenced more by instructions that encourage subjects to respond a certain way than those who score low on those metrics. In an attempt to be helpful during the FQs, subjects who are highly conscientious or who strongly desire social approval may utilize metamemory differently to rigorously follow instructions (e.g., respond frequently with "I don't know" in the IDK+ condition, resulting in accurate details being withheld) than subjects who are not conscientious or do not pursue social approval. Since this element of the study is exploratory and not part of previous research, I will not make predictions about the specific relationships between dependent variables and personality characteristics.

CHAPTER II

METHOD

Prior to data collection, all research materials and procedures were approved by the Institutional Review Board at the University of Louisville.

Design

The design of this study was a 2 (interview format: FN, FQs) × 5 (FQ instructions: Control, IDK+, IDK-, Accuracy+, Informative+) mixed factorial with the first factor being within-subjects and the second factor being between-subjects.

Subjects

Subjects were undergraduate students at the University of Louisville and participated during either the fall semester of 2017 or the spring semester of 2018. Subjects earned a one-hour credit in a psychology course for participating. Two hundred thirty-one subjects participated but responses from 28 subjects were unusable and not included in analyses. Subjects were removed if they skipped any FQs in less than seven seconds or they skipped the FN ($n = 21$), if they had previously seen the movie clip ($n = 2$), if they were younger than 18 ($n = 2$), or if the software collecting their responses crashed ($n = 3$). This left a total of $N = 203$ subjects ($M_{age} = 19.79$, $SD = 2.72$), most of whom were female (62.1%) and White (71.9%). Table 1 shows demographic information for each FQ instruction condition.

Table 1

Demographic information for each FQ instruction condition

FQ instruction condition	M_{age} (SD)	% Female	% White
Control ($n = 41$)	19.39 (2.72)	58.5	68.3
Accuracy+ ($n = 42$)	18.95 (1.19)	57.1	78.6
Informative+ ($n = 41$)	20.22 (2.91)	63.4	65.9
IDK+ ($n = 39$)	20.13 (2.56)	66.7	69.2
IDK- ($n = 40$)	20.30 (3.56)	65.0	77.5

Procedure

Prior to subjects arriving at the lab, the FQ instruction condition was determined based on which condition had the fewest subjects at that time. Up to five subjects could be run in each timeslot, and all subjects in each timeslot were assigned to the same condition. After providing informed consent, subjects began the study by watching the video described in the next section. The video was followed by a 10-minute break, during which subjects filled out two personality questionnaires. The order of questionnaires was counterbalanced. The remainder of the procedure was self-paced. First, subjects typed their FN. Next, subjects were instructed on how to answer the FQs (see Appendix A), and then provided their FQRs (see Appendix B), which were presented on their computer one at time in a randomized order. After subjects finished answering the FQs, they were asked, as a manipulation check, to type out any specific instructions they remembered reading before the FQs. They were also asked whether they have previously seen *The Brother's Bloom* (2008) (see Appendix C), and they were asked to provide their age, race, and sex. Figure 4 illustrates the procedure described above.
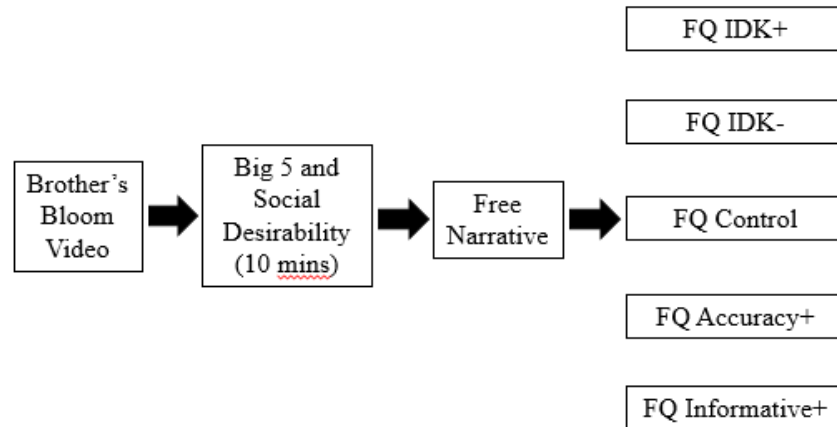
Figure 4. Illustration of the procedure, including all FQ conditions.

Materials

     Video. The video stimulus is a segment from the motion picture, *The Brother's Bloom* (2008). The segment depicts two brothers in the foster care system who perform a confidence game (con) in the town to which they recently moved. The con has several steps, and the segment shows both the written plan as well as the con being carried out by the brothers. The con involves the younger brother becoming friends with a group of children at the playground, including a girl he likes, in order to later be able to convince them to follow him to a cave. Once they get to the cave, the older brother tricks the children into running into the cave, where they fall and get their clothes dirty. After the brothers get in trouble, the con's payoff is revealed: the brothers had made a deal with the town's drycleaner to receive a cut of the profits they generated for him by getting the children's clothes dirty. The segment's duration is six minutes and 46 seconds. (https://www.youtube.com/watch?v=sUz0L4emNko)

     Big Five Inventory. The Big Five inventory (John & Srivastava, 1999) is a 44-item inventory that measures five personality factors. Subjects ranked each item on a 1 (*Disagree Strongly*) to 5 (*Agree Strongly*) scale. Items relevant to each personality factor

are randomly distributed, and several items must be reverse scored. The entire questionnaire was completed by subjects, but only their conscientiousness scores were analyzed.

Social Desirability Scale. The Crowne and Marlowe (1960) Social Desirability Scale consists of 33 items that measure a subject's need to obtain approval. Subjects responded True or False to each item, and a Social Desirability score was calculated by adding up the number of True responses to items that reflect social desirability and the number of False responses to items that reflect a lack of social desirability.

Coding Interviews

Two independent raters coded a subset (20%) of the FNs and FQRs for all dependent variables measured in each interview format (see Figure 3). In addition to the coding rubric and guidelines described below, all FNs and all FQRs were kept in separate documents and coded separately, so that knowledge of a subject's performance in one interview format did not influence coding of details in the other format. Subject numbers were the only identifying information for both FNs and FQRs in each document so that both raters were also blind to the instruction condition they were coding. The rater who coded only 20% of the FNs and FQs was also blind to the hypotheses. Interrater reliability of $r > .90$ was met, and the rest of the interviews were coded by a single rater. When new details were reported in the remaining interviews that did not have comparable details already in the coding rubric described below, the rater that coded all of the interviews continued to confer with the rater who was blind to the hypotheses.

In order to assist raters and minimize any chance of biased scoring, pilot testing was conducted to create a grading rubric (see Appendix D). During pilot testing, subjects

watched the Brothers Bloom video and were asked to write out general and specific details that could be used to answer the FQs. Subjects provided these details while watching the video (with the option to pause and revisit segments), not based on their memory. This generated many of the possible accurate details at varying precision levels, as well as using opposites for inaccurate details (e.g., the girl was accurately described as having "no freckles" during the pilot testing, therefore "having freckles" was added to the rubric as an inaccurate detail at the same precision level). However, some of the inaccurate details provided by subjects in the experiment were impossible to predict, therefore they were added post-hoc to the rubric after the raters consulted with each other. The coding rubric was based on FQ topics, thus FN details were only coded for the same topics.

Statements were judged as providing an "I don't know" response based on one of two requirements. When people responded with a statement that expressed uncertainty about their memory (e.g., "I'm not sure about…") or inability to provide any details (e.g. "I don't remember…"), and did not provide any details in addition to their "I don't know" statement, then it was counted as an "I don't know" response. However, when people expressed doubt in their memory, but still provided information (e.g., "I'm not sure how old she was, but I would guess six"), this was not counted as an "I don't know" statement as they still opted to report information. Sometimes "I don't know" responses were the only information provided in an FQR, but sometimes "I don't know" responses were provided regarding a specific aspect of their memory (e.g., "they got the candy from a candy store, but I don't know the name of the store"); in the latter situation, the part of the response that provided information was coded as accurate or inaccurate details in

addition to an "I don't know" statement. The majority of "I don't know" statements were identified through these guidelines, but it is possible that subjects could have simply opted to skip the question after reading it, and thus still opted to withhold information. After timing how long it took a person unfamiliar with the questions to read through each question, it was decided that if an FQ was skipped after being on the page for seven seconds, the skip would be counted as an "I don't know" response. When a subject skipped the FQ in less than seven seconds, it was counted as a mistakenly skipped question, and the responses was not scored or analyzed.

CHAPTER III

RESULTS

FQ All

This section comprises analyses of the details provided in response to all six FQs. Details provided in response to all FQs were included, regardless of whether corresponding topics were mentioned in the preceding FN. The same nine dependent variables were measured in both interview formats and each of these was submitted to a 2 (interview format: FN, FQRs) × 5 (FQ instructions: Control, Accuracy+, Informative+, IDK+, IDK-) mixed-design ANOVA, with interview format being the within-subjects factor and FQ instructions being the between-subjects factor. "I don't know" responses were counted in FQRs only. Counts were submitted to a one-way between-subjects ANOVA. Social desirability and conscientiousness were correlated with all of the FQR dependent variables.

As previously noted in the Method chapter, not all FN details were included in the analyses. Only FQ-topic details were coded for in the FNs. This should be noted, as the number of details coded for in the FNs ($M = 3.10$, $SD = 4.02$) was dramatically different compared to the number of words in subjects' FNs ($M = 91.03$, $SD = 72.42$).

Report Option

Quantity of Accurate Details. There was a significant main effect of interview format, $F(1, 198) = 627.69$, $p < .001$, $\eta_p^2 = .760$, with subjects providing more details in FQRs ($M = 19.38$) than FNs ($M = 2.48$). $F(1, 198) = 627.69$, $p < .001$, $\eta_p^2 = .760$. Neither the main effect of FQ instructions, $F(4, 198) = 2.03$, $p = .092$, $\eta_p^2 = .039$, nor the interaction, $F(4, 198) = 1.10$, $p = .360$, $\eta_p^2 = .022$ were significant.

Figure 5. Mean number of accurate details as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Quantity of Inaccurate Details. The main effect of interview format, $F(1, 198) = 222.61$, $p < .001$, $\eta_p^2 = .563$, and the main effect of FQ instructions, $F(4, 198) = 2.80$, $p = .027$, $\eta_p^2 = .054$, were both significant. However, these effects were qualified by a significant interaction, $F(4, 198) = 2.80$, $p = .027$, $\eta_p^2 = .054$. To decompose this interaction, separate one-way ANOVAs were conducted on both interview formats. For FNs, the effect of FQ instructions was not significant, $F(4, 198) = .626$, $p = .645$, $\eta_p^2 = .012$, which was expected because all subjects received the same instructions for FNs.

For FQRs, the effect of FQ instructions was significant, $F(4, 198) = 57.05$, $p = .019$, $\eta_p^2 = .058$. To determine which FQ instructions led to significant differences, independent t tests were conducted. Subjects in the IDK+ condition provided significantly fewer inaccurate details ($M = 3.62$) than subjects in the Control ($M = 6.80$) and IDK- ($M = 5.98$) conditions, smallest $t(77) = 2.84$, $p = .006$, $d = .638$, which did not differ from each other, $t(79) = .82$, $p = .414$, $d = .183$. Subjects in the Accuracy+ ($M = 4.93$) and Informative+ ($M = 5.17$) conditions did not differ from subjects in any other condition, smallest $t(81) = 1.91$, $p = .059$, $d = .420$.
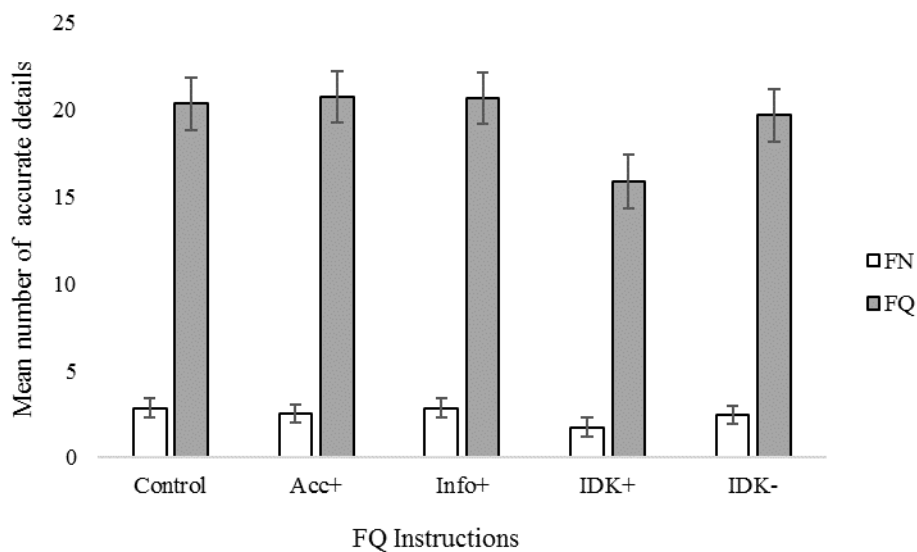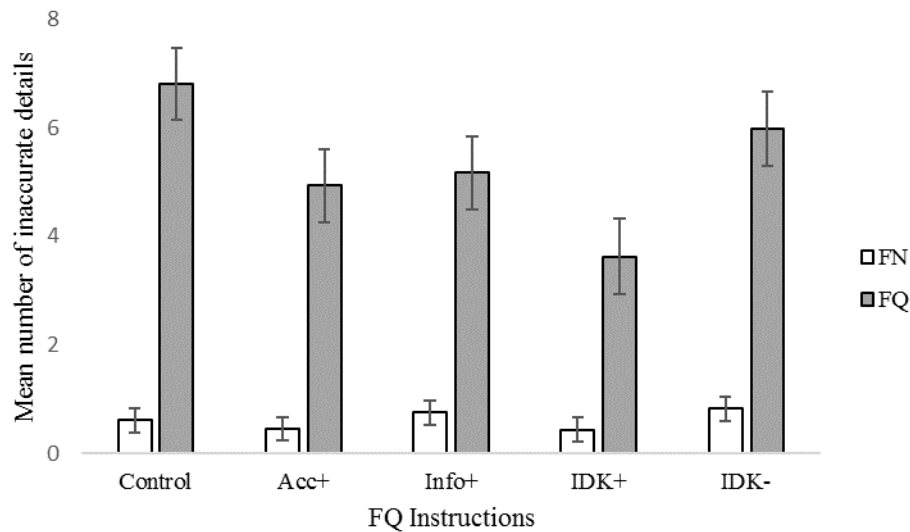

Figure 6. Mean number of inaccurate details as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Quantity of Incorrect-category inaccurate detail. There was a significant main effect of interview format, $F(1, 198) = 203.42$, $p < .001$, $\eta_p^2 = .507$, with subjects providing more details during FQRs ($M = 3.61$) then FNs ($M = .40$). Neither the main effect of FQ instructions, $F(4, 198) = 2.20$, $p = .071$, $\eta_p^2 = .042$, nor the interaction, $F(4, 198) = 1.62$, $p = .170$, $\eta_p^2 = .032$, were significant.
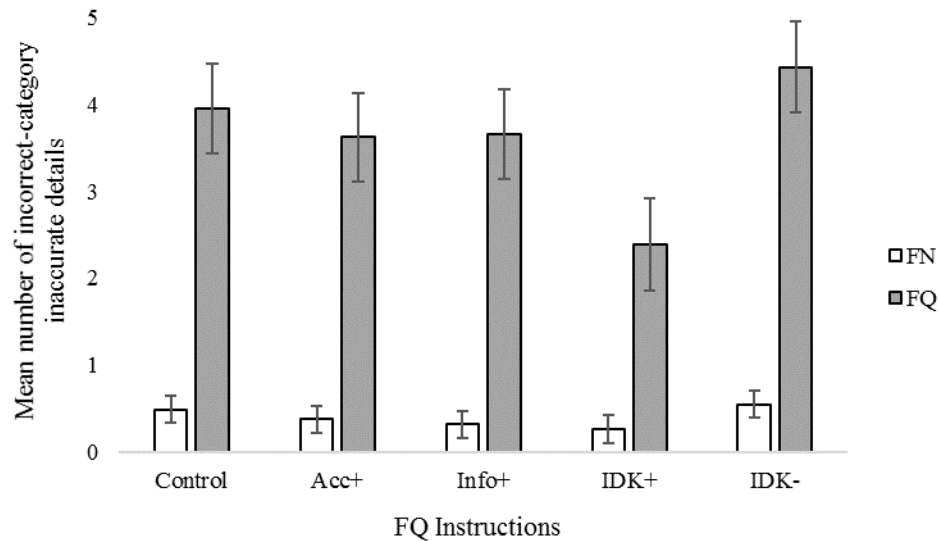
Figure 7. Mean number of incorrect-category inaccurate details as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Number of "I don't know" responses. FQ instructions had a significant effect on the number of "I don't know" responses, $F(4, 198) = 3.24$, $p = .013$, $\eta^2 = .061$. To determine which FQ instructions caused subjects to use "I don't know" responses differently, independent *t* tests were conducted. Subjects in the IDK+ condition opted to utilize "I don't know" responses significantly more frequently ($M = 1.33$) than subjects in all other conditions (.58 ≤ *Ms* ≤ .86), smallest $t(79) = 2.13$, $p = .036$, $d = .474$. Subjects in the Control, Accuracy+, Informative+, and IDK- conditions did not significantly differ from each other, largest $t(80) = 1.46$, $p = .147$, $d = .323$.

To better understand whether subjects were saying "I don't know" to avoid providing inaccurate details, without also withholding accurate details, correlations between frequency of "I don't know" responses and inaccurate and accurate details were conducted in each FQ condition. There was a significant negative relationship between frequency of "I don't know" responses and number of accurate details in the Accuracy+ and Informative+ conditions, smallest $r = -.357$, $p = .022$. The IDK+ and Control

conditions both had a significant negative relationship between frequency of "I don't know responses and number of inaccurate details, smallest $r = -.341$, $p = .029$. The combination of "I don't know" responses *having* a relationship with inaccurate details, but *not* with accurate details indicates people are effectively opting to use "I don't know" responses to filter out inaccurate details, without also withholding accurate details. The IDK+ and Control conditions were the only conditions that resulted in this specific combination.
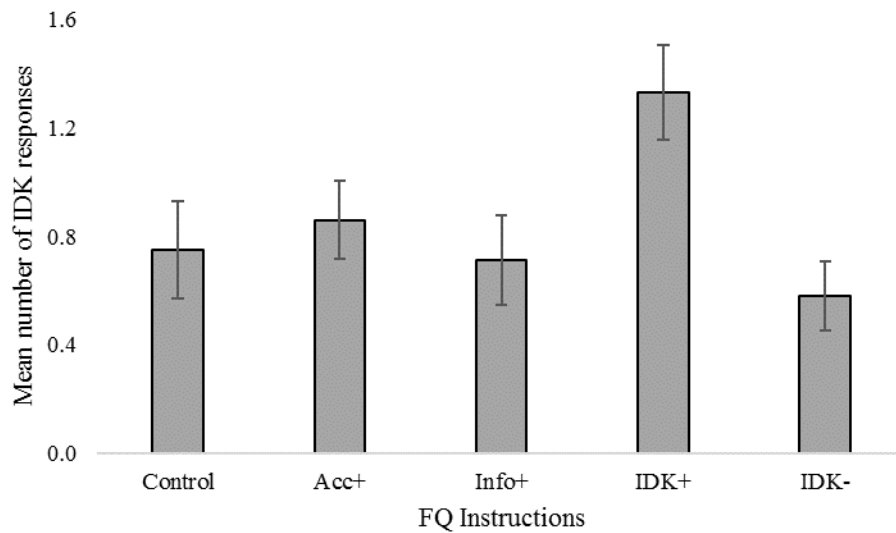
Figure 8. Mean number of "I don't know" responses as a function of FQ instructions. Error bars represent ± *SEM*.

Grain Size

Precision of Accurate Details. Precision scores ranged from 1 (*General*) to 5 (*Specific*). There were fewer subjects in each condition (*n*s: 18-24) for the analysis of accurate detail precision than in the quantity analyses, because responses that contained no details could not be scored for precision. There was a significant main effect of interview format, $F(1, 98) = 25.13$, $p < .001$, $\eta_p^2 = .204$, with subjects providing more precise details during FQR (*M* = 3.66) than during FNs (*M* = 3.16). Neither the main

effect of FQ instructions, $F(4, 98) = 1.41$, $p = .237$, $\eta_p^2 = .054$, nor the interaction, $F(4, 98) = 1.48$, $p = .213$, $\eta_p^2 = .057$, were significant. As previously mentioned, group sizes were reduced due to the limited number of subjects who provided accurate details during FNs. In order to more fully understand the influence of FQ instructions on the precision level of accurate details during FQRs, an additional one-way ANOVA was conducted, using all of the subjects' data from FQRs. This analysis confirmed the previous finding that the main effect of FQ condition was not significant, $F(4, 198) = .54$, $p = .705$, $\eta_p^2 = .011$.
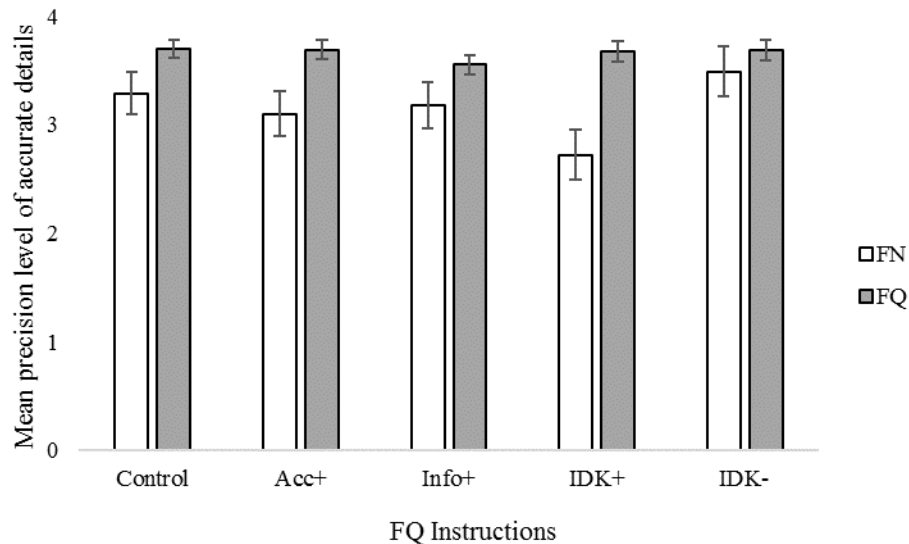
Figure 9. Mean precision level of accurate details as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Precision of Inaccurate Details. The analyses on precision level of inaccurate details was based on fewer than 15 people per condition, and thus will not be reported. However, a one-way ANOVA was conducted using all of the subjects' FQR data. The main effect of FQ instructions was not significant, $F(4, 180) = .93$, $p = .450$, $\eta_p^2 = .020$.
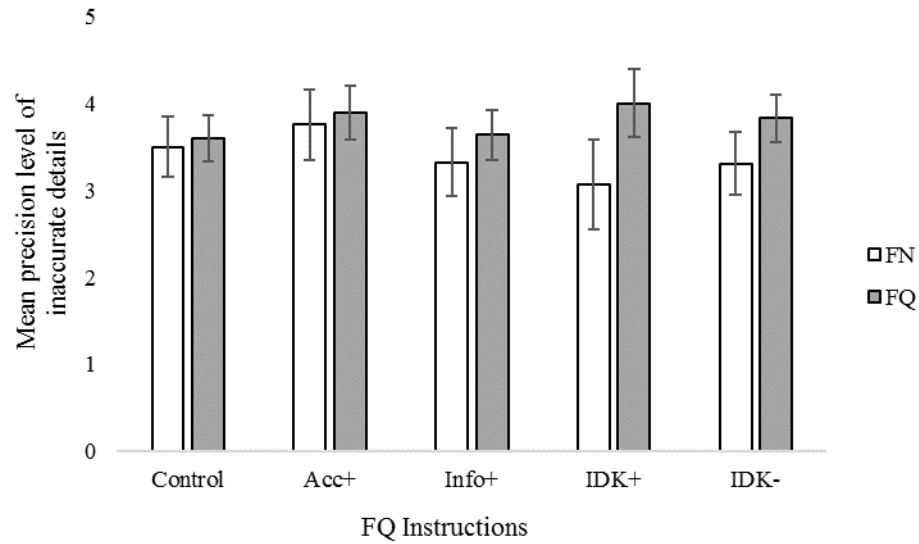
Figure 10. Mean precision level of inaccurate details as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Quantity of Correct-category inaccurate details. There were significant main effects of both interview format, $F(1, 198) = 87.20$, $p < .001$, $\eta_p^2 = .306$, and FQ instructions, $F(4, 198) = 2.52$, $p = .043$, $\eta_p^2 = .048$, but these were qualified by a significant interaction, $F(4, 198) = 4.19$, $p = .003$, $\eta_p^2 = .078$. To explore this interaction, a one-way ANOVA was conducted for each interview format. As would be expected, the effect of FQ instructions was significant for FQRs, $F(4, 198) = 3.55$, $p = .008$, $\eta_p^2 = .067$, but not FNs, $F(4, 198) = 1.09$, $p = .360$, $\eta_p^2 = .022$. Independent *t* tests comparing the number of correct-category inaccurate details between FQ instructions were conducted after visual inspection of Figure 11, which showed that subjects in the Control condition reported a larger number of correct-category inaccurate details during FQRs than subjects in all other conditions, smallest $t(79) = 2.20$, $p = .030$, $d = .490$.
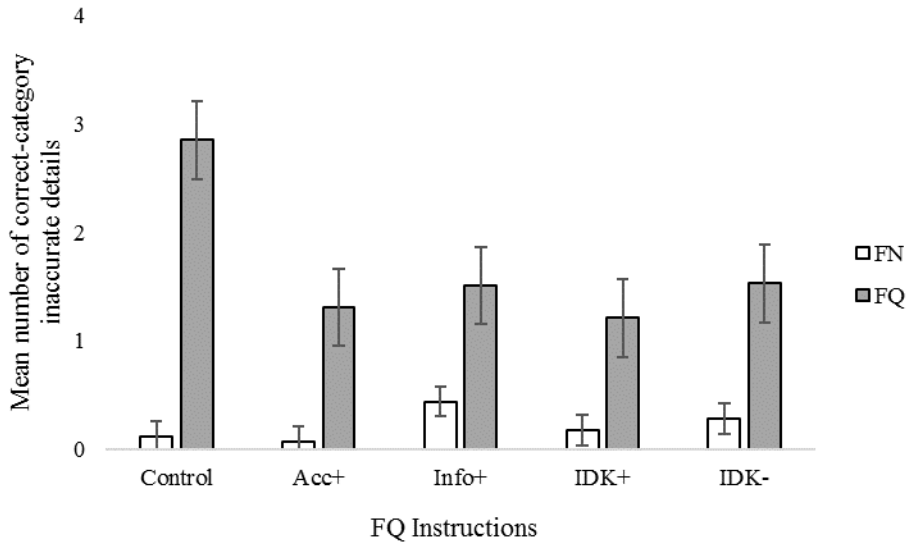
Figure 11. Mean number of correct-category inaccurate details as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Tradeoffs

Overall Accuracy. As with precision, when subjects did not provide any information, accuracy could not be calculated. Thus, the number of subjects that could be included in this analysis (*n*s: 19-26) was lower than for the quantity analyses because few people provided useable details during their FNs. The main effect of interview format, $F(1, 107) = .69$, $p = .407$, $\eta_p^2 = .006$, the main effect of FQ instructions, $F(4, 107) = .26$, $p = .902$, $\eta_p^2 = .010$, and the interaction, $F(4, 107) = 1.43$, $p = .228$, $\eta_p^2 = .051$, were all not significant. In order to better understand the influence of FQ instructions on the overall accuracy of FQRs without restricting analysis to subjects who provided details in their FNs, an additional one-way ANOVA was conducting using all of the subjects' data from their FQRs. This analysis confirmed the previous finding that the FQ instructions did not affect overall accuracy of FQRs, $F(4, 198) = 1.93$, $p = .107$, $\eta_p^2 = .038$.
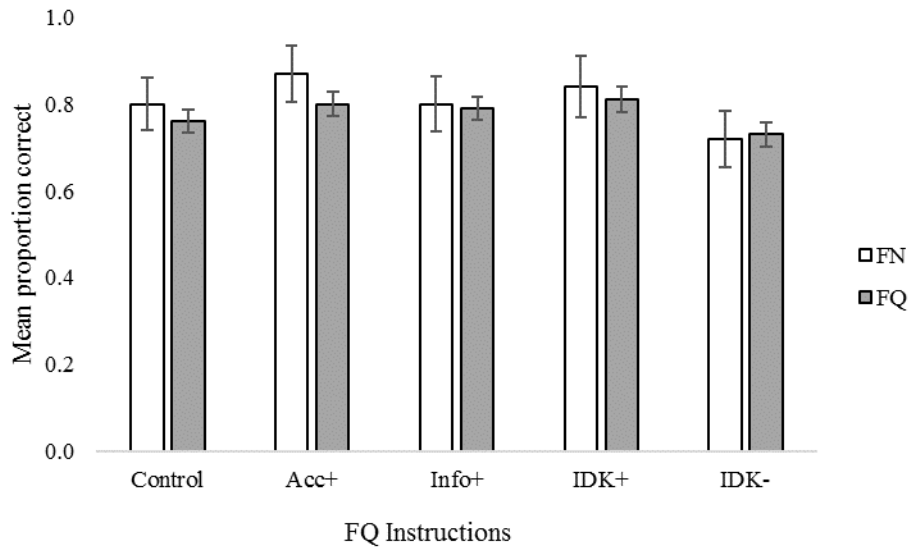
Figure 12. Mean proportion correct as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Total Quantity. The main effect of interview format was significant, $F(1, 198) = 756.76$, $p < .001$, $\eta_p^2 = .793$, with subjects providing more details in FQRs ($M = 24.68$) than in FNs ($M = 3.10$). The main effect of FQ instructions was also significant, $F(4, 198) = 2.82$, $p = .026$, $\eta_p^2 = .054$, with subjects in the IDK+ condition ($M = 21.67$) providing fewer details than subjects in all other conditions ($28.67 \leq Ms \leq 30.61$), smallest $t(79) = 2.62$, $p = .011$, $d = .58$). The interaction was not significant, $F(4, 198) = 1.95$, $p = .103$, $\eta_p^2 = .038$.
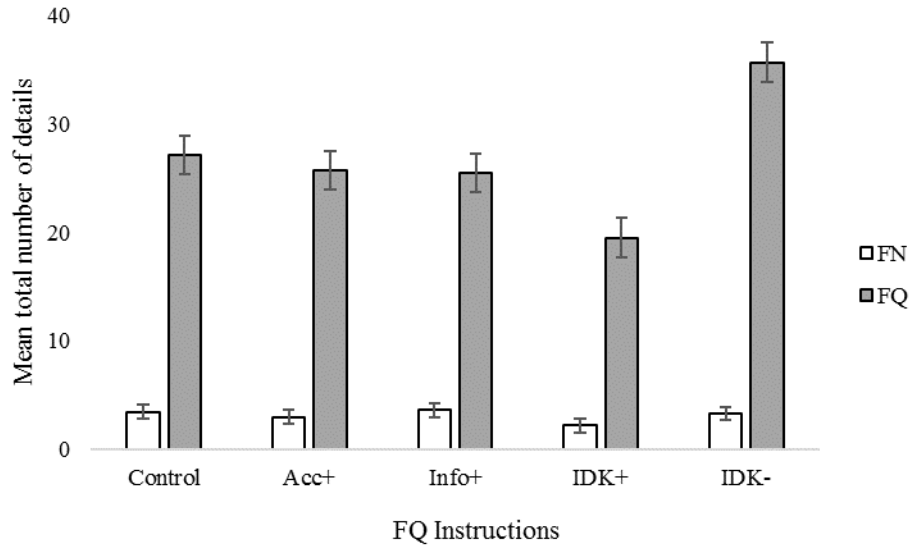
Figure 13. Mean total number of details as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Overall Precision. As with the other precision analyses reported, some subjects had missing data due to not reporting any useable details. Thus, the number of subjects that could be included in this analysis (*n*s: 19-26) was lower than for the quantity analyses because few people provided inaccurate details during their FNs. The main effect of interview format was significant, $F(1, 107) = 38.14$, $p < .001$, $\eta_p^2 = .263$, with subjects reporting more precise information in FQRs ($M = 3.68$) than in FNs ($M = 3.15$). Neither the main effect of FQ instructions, $F(4, 107) = 1.56$, $p = .191$, $\eta_p^2 = .055$, nor the interaction, $F(4, 107) = 2.06$, $p = .091$, $\eta_p^2 = .071$, were significant. In order to better understand the influence of FQ instructions on overall precision of FQRs without restricting the data to subjects who provided details in their FNs, an additional one-way ANOVA was conducting using all of the subjects' data from FQRs. This analysis confirmed the previous finding that the FQ instructions did not affect overall precision of FQRs, $F(4, 198) = .64$, $p = .636$, $\eta_p^2 = .013$.
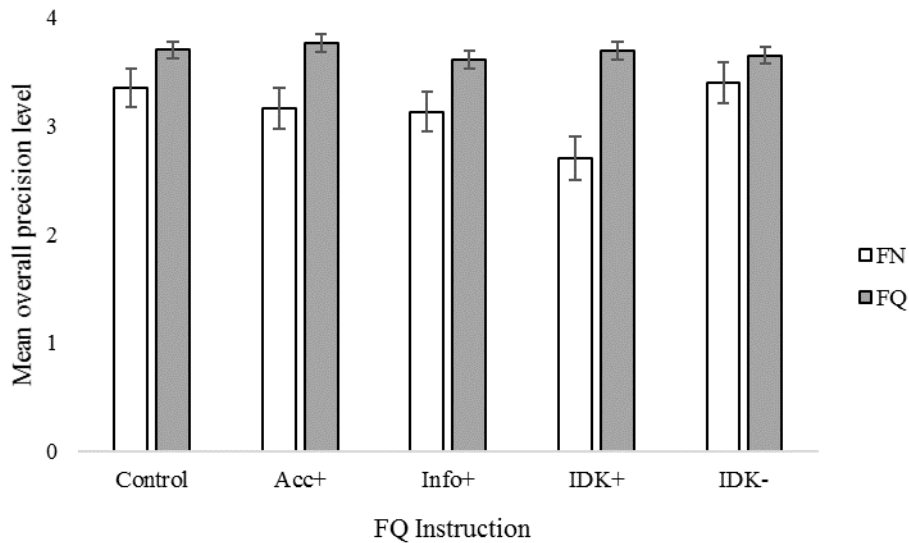
Figure 14. Mean overall precision level as a function of interview format and FQ instructions. Error bars represent ± *SEM*.

Individual Differences

The following tables provide correlations between the personality characteristics of interest, social desirability scores and conscientiousness, and the various FQR dependent measures. In the Control condition, social desirability scores had a significant negative relationship with quantity of inaccurate details, correct category and incorrect category inaccurate details, and total quantity of details, as well as a significant positive relationship with overall FQR accuracy. In the Accuracy+ condition, social desirability scores had a significant positive relationship with precision of accurate details and overall precision level of FQRs. In the Informative+ condition, social desirability scores and conscientiousness had a significant positive relationship with the number of "I don't know" responses given. Notably, social desirability was correlated with more dependent measures than would be expected by chance, but conscientiousness was not. There were

no significant correlations between personality measures and FQRs in the IDK+ or IDK-

conditions.

Table 2

Correlations between Social Desirability scores and FQR variables

FQ Instruction	Accurate Quantity	Accurate Precision	Inaccurate Quantity	Inaccurate Precision	Correct-Category Inaccurate Detail	Incorrect-Category Inaccurate Detail	Number of "IDK" responses	Overall Accuracy	Overall Precision	Overall Quantity
Control	-.16	-.06	-.49**	.19	-.42*	-.34*	.16	.45*	-.05	-.32*
Accuracy+	.13	.35*	.20	-.04	-.02	.22	-.05	-.17	.34*	.18
Informative+	.02	-.19	-.12	-.13	-.03	-.15	.44*	.06	-.15	-.04
IDK+	.04	-.30	-.08	.02	.01	-.12	.17	.15	-.27	.00
IDK-	-.03	-.07	-.06	-.06	-.09	-.05	-.08	-.07	-.14	-.05

*p < .05
**p < .001

Table 3

Correlations between Conscientiousness scores and FQR variables

FQ Instruction	Accurate Quantity	Accurate Precision	Inaccurate Quantity	Inaccurate Precision	Correct-Category Inaccurate Detail	Incorrect-Category Inaccurate Detail	Number of "I don't know" responses	Overall Accuracy	Overall Precision	Total Quantity
Control	-.22	-.05	-.20	.21	-.05	-.24	.15	.15	-.01	-.26
Accuracy+	.10	.20	-.13	.13	.04	-.16	.22	.00	.21	.05
Informative+	.08	-.14	-.19	-.04	-.18	-.15	.35*	.27	-.11	-.02
IDK+	.15	-.14	.03	-.14	.16	-.09	-.20	-.05	-.09	.14
IDK-	.05	-.03	-.14	.00	-.09	-.13	-.04	.13	-.07	.00

*p < .05

**p < .001

Manipulation Check

After completing the full interview, subjects were asked to report details that they remembered from the instructions preceding the FQs (see Appendix C). While subjects received additional instructions other than those related to metamemory (e.g., which key will move them to the next FQ), reported instructions related to metamemory were the only details checked for accuracy. Of the subjects in each experimental condition, 5.0% in the Accuracy+ condition, 17.9% in the IDK+ condition, 29.2% in the Informative condition, and 40.0% in the IDK- condition mentioned the metamemory instructions they received prior to the FQs during the manipulation check.

FQ Mentions

The same analyses were run a second time, but this time FQR details were only included if corresponding FQ topics were mentioned in the preceding FN. Excluding subjects who did not mention any FQ topics in their FN reduced the number of subjects in each condition (*n*s: 26-32). Similar to FQ All analyses, some subjects had missing data for precision analyses, which further reduced the number of subjects in each condition.

The same nine dependent variables were submitted to a 2 (interview format: FN, FQRs) × 5 (FQ instructions: Control, Accuracy+, Informative+, IDK+, IDK-) mixed ANCOVA, with interview format being the within-subjects factor and FQ instructions being the between-subjects factor and the number of FQ topics mentioned in a subjects' FN as the mean-centered covariate (Delaney & Maxwell, 1981). "I don't know" responses were counted in FQRs only. Counts were submitted to a one-way between-

subjects ANCOVA with the same covariate. Social desirability and conscientiousness were correlated with all of the FQR dependent variables.

Results of analyses on seven variables (quantity of accurate details, quantity of incorrect-category inaccurate details, precision of accurate details, quantity of correct-category inaccurate details, overall accuracy, overall precision, and total quantity) were the same when restricted to FQ Mentions as when using FQ All. The analyses on precision level of inaccurate details was based on fewer than 15 people per condition, and thus will not be reported. Additionally, the FQ mentions data showed the same association between FQ instructions and whether the metamemory instructions were mentioned in the manipulation check. Results of FQ Mention analyses that *differed* from FQ All analyses are reported below.

Report Option

Quantity of Inaccurate Details. There was a significant main effect of interview format, $F(1, 144) = 47.94$, $p < .001$, $\eta_p^2 = .250$, with subjects providing more inaccurate details during FQRs ($M = 2.58$) then during FNs ($M = .827$). The main effect of FQ instructions was not significant, $F(4, 144) = .74$, $p = .569$, $\eta_p^2 = .020$. Unlike with the FQ All analyses, the interaction was not significant, $F(4, 144) = .68$, $p = .610$, $\eta_p^2 = .018$.
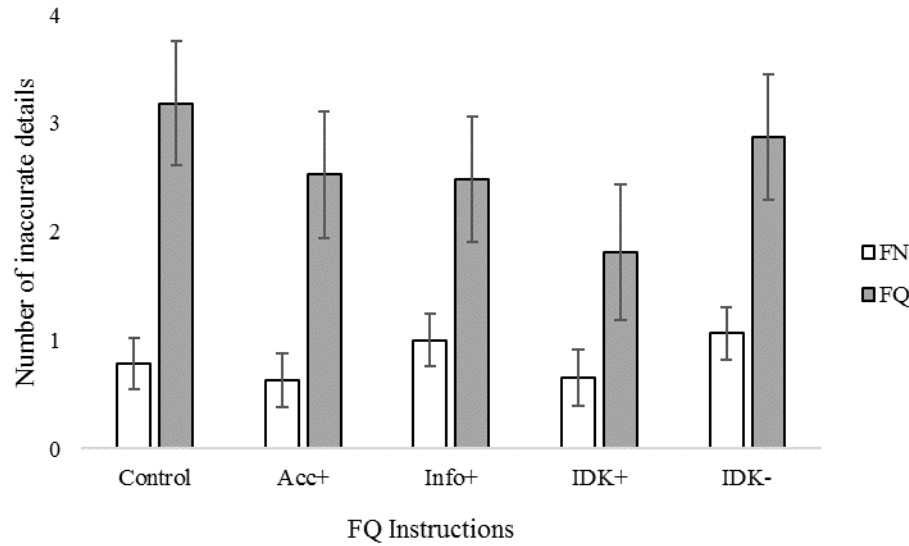
Figure 15. Mean number of inaccurate details as a function of interview format and FQ instructions based on FQ Mentions. Error bars represent ± *SEM*.

Number of "I don't know" responses. When using all subjects who mentioned at least one FQ topic in their FNs, Levene's test was significant, $F(4, 145) = 3.41$, $p = .011$, therefore the data was trimmed so that group sizes would be equal ($ns = 26$). Unlike with the FQ All analyses, there was no differences in use of "I don't know" responses between FQ instructions, $F(4, 124) = .70$, $p = .597$, $\eta_p^2 = .022$. Across all FQ conditions, there were no significant correlations between inaccurate or accurate details and frequency of "I don't know" responses.
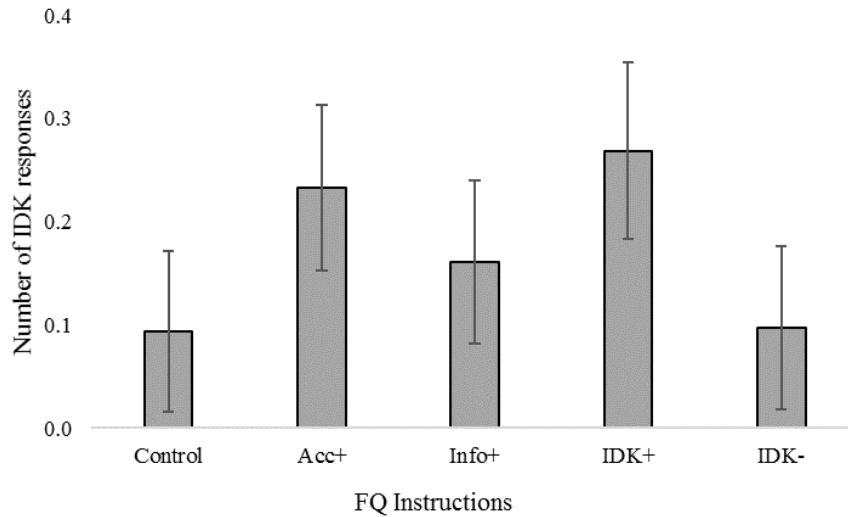
Figure 16. Mean number of "I don't know" responses as a function of FQ instructions based on FQ Mentions. Error bars represent ± *SEM*.

Individual Differences

The following tables provide the correlations between scores measuring the personality characteristics of interest, social desirability and conscientiousness, and the FQR dependent variables. The positive relationship between social desirability and overall accuracy in the Control condition remained significant. No other significant correlations from the FQ All remained significant when limited to FQ Mentions. In the IDK+ condition, social desirability had a significant negative relationship with overall precision of FQRs and conscientiousness had a significant negative relationship with precision of accurate details, neither of which were significant in the FQ All analyses.

Table 4

Correlations between Social Desirability scores and FQR variables

FQ Instruction	Accurate Quantity	Accurate Precision	Inaccurate Quantity	Inaccurate Precision	Correct-Category Inaccurate Detail	Incorrect-Category Inaccurate Detail	Number of "IDK" responses	Overall Accuracy	Overall Precision	Overall Quantity
Control	.20	-.08	-.19	.19	-.31	.04	.04	.36*	-.27	-.10
Accuracy+	.17	.04	.18	-.44	-.01	.21	.06	-.18	.03	.19
Informative+	.01	.17	-.10	-.31	-.07	-.11	.30	.07	.20	-.03
IDK+	.06	-.17	-.11	.15	-.09	-.08	-.03	.19	-.43*	.01
IDK-	.00	.02	-.04	-.26	-.14	.04	.33	.02	-.18	-.01

*p < .05
**p < .001

46

Table 5

Correlations between Conscientiousness scores and FQR variables

FQ Instruction	Accurate Quantity	Accurate Precision	Inaccurate Quantity	Inaccurate Precision	Correct-Category Inaccurate Detail	Incorrect-Category Inaccurate Detail	Number of "I don't know" responses	Overall Accuracy	Overall Precision	Total Quantity
Control	-.08	-.26	-.09	-.35	-.18	-.05	.29	.04	-.26	-.09
Accuracy+	-.02	-.28	-.12	.13	.32	-.30	.07	.09	-.20	.06
Informative+	.20	-.01	.06	-.19	.00	.09	.27	-.21	.02	.16
IDK+	-.07	-.55**	.02	.30	.25	-.28	-.23	-.14	-.29	-.04
IDK-	-.09	.23	-.29	-.47	-.36	-.12	.06	.21	.06	-.17

*$p < .05$
**$p < .001$

CHAPTER IV

DISCUSSION

The goals of this dissertation were to examine the effects of interview format and of FQ instructions on quality of FQRs, and explore the relationship between personality characteristics and FQR quality. Previous research using the standard interview format (FN followed by FQs) has consistently shown that format has an effect on the accuracy of details (e.g., Evans & Fisher, 2011; Hilgard & Loftus, 1979; Lipton, 1979; Odinot et al., 2009; Taylor & Dando, 2018; Yarmey & Yarmey, 1997). To date, though, no study has examined *why* this occurs. The criminal justice system could benefit not only from understanding why accuracy is relatively low in FQRs, but also whether FQR accuracy can be improved.

Research examining how people monitor their memories (i.e., metamemory) indicates that accuracy of details often has an inverse relationship with quantity and precision of those details. When people are able to choose whether to report or withhold a detail (i.e. report option) they provide fewer—but more accurate—details than when they are required to answer every question (Korat & Goldsmith, 1994, 1996; Koriat et al., 2001), resulting in an accuracy-quantity tradeoff. Likewise, when people are able to control how precise or general (i.e., grain size) their answers are, they provide less precise, but more accurate, details than when they are required to provide only exact

answers (Ackerman & Goldsmith, 2008; Goldsmith et al., 2005; Koriat et al., 2002), resulting in an accuracy-precision tradeoff. Therefore, accuracy, quantity, and precision were measured in this dissertation to better understand how people utilize their metamemory.

Contrary to the findings of previous studies (e.g., Evans & Fisher, 2011; Hilgard & Loftus, 1979; Lipton, 1979; Odinot et al., 2009; Taylor & Dando, 2018; Yarmey & Yarmey, 1997), and my expectations, FQRs in this study were just as accurate as FNs, on average. While the FQRs in this study were roughly as accurate as expected, the FNs were less accurate ($M = .80$ here, compared to .90 or higher in prior studies). Procedural differences between this study and previous studies might explain the lower FN accuracy rate, so these differences will be reviewed in the theoretical implications discussed below. These results are poorly suited to improve our understanding of the role that metamemory plays in the accuracy-related effects of interview format, and may in fact indicate that accuracy-related effects of interview format are not inevitable. Furthermore, because accuracy was equivalent across interview formats, any accuracy-quantity or accuracy-precision tradeoffs are impossible to discern. However, the results can still strengthen our understanding of quality of details reported in different interview formats and in response to different FQ instructions.

The quantities of accurate-, inaccurate-, and incorrect-category inaccurate details were analyzed because of their potential to show how people utilize their report option. Metamemory research indicates that the availability of report option is closely linked to the quantity and accuracy of details. Incorrect-category inaccurate details have not been measured in previous research, but they indicate that an inaccurate detail is not even in

the correct general category (e.g., reporting a dark-colored car as white). This type of detail should be withheld (i.e., utilize report option) by subjects who are trying to remain accurate.

The level of precision of accurate and inaccurate details and the number of correct-category inaccurate details were measured as indicators of how people were using their control over detail grain size. Metamemory research shows that respondents' ability to control the grain size of details is related to the accuracy and specificity of their answers. Correct-category inaccurate details identify inaccurate details that were reported at a level of detail that was too precise (e.g., reporting a black car as blue, which are both dark colors), as opposed to the previously discussed incorrect-category inaccurate details. Inaccurate details have not previously been categorized as correct- or incorrect-category details in the literature, but this approach allows inaccurate details to be attributed to the report option (incorrect category) or grain size (correct category).

Effect of Interview Format

All of the variables related to report option were higher in the FQRs than in the FNs when no metamemory instructions were provided prior to FQs, supporting my hypothesis regarding quantity of inaccurate details, but not my hypothesis regarding quantity of accurate details. Collectively, this may indicate that people may have a lower accuracy criterion when responding to FQs than when providing FNs. A lower criterion during the FQs would allow more details of all kinds to be reported and would fail to filter out as many inaccurate details. This pattern *might* suggest that people naturally utilize their report option differently depending on the interview format. However, the lack of a difference in detail accuracy between FNs and FQRs does not support this

50

interpretation, because experimental manipulation of report option in the metamemory literature typically leads to an accuracy-quantity tradeoff (Koriat & Goldsmith, 1994, 1994; Koriat et al., 2001). Alternate explanations for the effects of interview format on detail quality will be discussed in the theoretical implications section below.

Unfortunately, only the precision of accurate details and number of correct-category inaccurate details could be analyzed to examine the effect of interview format on precision level. The analysis of the precision of inaccurate details was based on a small number of subjects and the results could not be interpreted, so they are not reported. The precision of accurate details was higher (responses were more specific) in FQRs than in FNs, which supported my hypothesis, and there were more correct-category inaccurate details reported in FQRs than in FNs when no metamemory instructions were provided. This could reflect a change in how subjects utilized their control over grain size in FNs compared to FQRs, but, again, if this were correct, the results should have included an effect of interview format on accuracy to match the accuracy-precision tradeoff common in the metamemory literature. Thus, this demonstrates that FQRs can elicit more precise details without necessarily eliciting less accurate details.

Effect of FQ Instructions

Subjects appear to be capable of intentionally filtering out inaccurate details in FQRs when they are encouraged to respond with "I don't know," as needed. When given IDK+ instructions, subjects said "I don't know" more often than when they were given other metamemory instructions (Accuracy+, Informative+, or IDK-), or when they were given no metamemory FQ instructions (Control), which supported my hypothesis. In the IDK+ and Control conditions, the frequency of "I don't know" responses was negatively

related to the number of inaccurate details, but was not related to the number of accurate details. Importantly, subjects in the IDK+ condition reported significantly fewer inaccurate details than subjects in the Control condition—probably because IDK+ subjects opted to say "I don't know" more frequently—but subjects in the IDK+ condition did not differ significantly from subjects in other conditions in the number of accurate details reported, which supported my hypotheses. This builds on previous evidence that people are more likely to say "I don't know" and report fewer erroneous details during FQs when they are encouraged to use "I don't know" statements than when they are not (Scoboria & Fisico, 2013). Subjects in the Accuracy+, Informative+, and IDK- conditions did not provide different quantities of details (accurate or inaccurate) compared to each other or to those in the Control condition, which did not support my hypotheses.

While the difference was not statistically significant, subjects in the IDK+ condition did report numerically fewer accurate details than those in other groups. Due to this decrease, the significantly lower number of inaccurate FQR details that were elicited by the IDK+ condition did not result in a significantly higher accuracy rate, relative to the other FQ conditions, which did not support my hypothesis. Additionally, the total quantity of details provided in both interview formats was significantly lower in the IDK+ condition than in the other FQ conditions. Together, these findings indicate that more research is needed to determine whether IDK+ instructions selectively reduce inaccurate details while preserving accurate details, or whether they simply reduce the reporting of details across the board.

In general, it seems that FQ instructions have little to no effect on precision level of details, which did not support my hypothesis. The precision levels of accurate FQR details were equivalent across conditions, as were precision levels of inaccurate details. However, subjects in the Control condition provided more correct-category inaccurate details than those in any other condition. This could indicate that including *any* metamemory instructions prior to FQ could cause people to be more cautious of reporting details at a more specific level than they are confident in, yet, the collective evidence indicates that precision of details was the same across groups. Thus, there is little reason to believe that metamemory instructions affect how people utilize their control over grain size in FQRs.

Individual Differences

I also examined the relationship between personality characteristics and FQR quality based on FQ instructions received. The results indicate that subjects who pursue social approval are more likely to defy (explicit or assumed) instructions, which is counterintuitive to what one might expect. In the absence of metamemory instructions (Control condition), based on previous findings of more details provided during FQRs at the cost of accuracy, I expected subjects to assume that additional questions were being asked as a cue to report more details (even if it came at a cost to accuracy). Yet subjects who highly desire social approval maximized accuracy by reporting fewer inaccurate details. When subjects were encouraged to prioritize accuracy during FQRs, I hypothesized that they would avoid reporting precise details in order to increase their accuracy, yet subjects who scored high on the social desirability scale provided accurate details that were highly precise. Instructions that encourage subjects to prioritize

informativeness were expected to cause people to avoid saying "I don't know" in order to increase the number of details they report. However, I found that subjects who value social approval, as well as those who were highly conscientious, were more like to give "I don't know" responses than other subjects.

Unlike in the Control, Accuracy+, and Informative+ conditions there was no relationship between personality and FQR quality in the IDK+ or IDK- conditions. While this is surprising, because people who are highly conscientious or have a high need for social desirability might have followed directions to an extreme (e.g., use "I don't know" too much or too little, respectively), it is encouraging that there were not significant individual differences in the quality of FQRs in the IDK+ condition related to these personality characteristics. The IDK+ condition was the most effective at statistically reducing the quantity of inaccurate details without statistically reducing the quantity of accurate details. The potential utility of the IDK+ approach is bolstered by the observation that subjects who scored low on social desirability still appeared to follow instructions, while those who scored high on social desirability did not use the "I don't know" so frequently that they reported significantly fewer accurate details.

The difference between explicit and implicit expectations communicated by metamemory instructions might explain why there were individual differences in detail quality in response to some FQ instructions, but not others. Accuracy+ and Informative+ instructions were explicit about what subjects should prioritize, but they did not instruct subjects on how to respond, whereas IDK+ and IDK- instructions were explicit about how subjects should respond, but they did not tell subjects what to prioritize. The Control instructions did neither, leaving subjects to make assumptions regarding what to

prioritize and how to respond. Perhaps because IDK+ and IDK- instructions were explicit in terms of *how* to answer, everyone was able to follow the guidelines equally well. In contrast, the Accuracy+, Informative+, and Control instructions left *how* to answer questions more ambiguous, so individual differences may have been more likely to influence how a person attempted to meet the goal of the instructions.

FQ Mention Analyses

Effects of how FQs are chosen on the quality of FQRs—and thus the difference in quality of FNs and FQRs—are not clear because previous studies have differed in how FQs were determined. Some researchers have used trained interviewers to follow up on topics mentioned in a person's FN (e.g., Memon et al., 1997; Odinot et al., 2009; Taylor & Dando, 2018), while others used a predetermined set of FQ questions that were asked regardless of topics covered in a subject's FN (e.g., Evans & Fisher, 2011; Lipton, 1977). Thus, in addition to analyzing the data based on all of the FQRs, I repeated my analyses using only FQRs that corresponded to FQ topics had been mentioned in the preceding FN. This approach revealed a few systematic differences.

Unlike when using all FQR details, none of the FQ instructions seemed to have an effect on quality of details. Notably, there was no difference in the rate of "I don't know" responses nor the number of inaccurate details across FQ instructions. These results are reasonable, as people are likely to be able to provide information about topics that they brought up on their own, even if it is just repeating the information provided in their FNs. However, it should be noted that the analyses on both the rate of "I don't know" responses and number of inaccurate details, like many of the FQ Mention analyses, were underpowered (highest power = .23).

Many of the relationships between social desirability and conscientiousness scores and FQR quality were no longer evident when data was restricted to FQ Mentions, suggesting that many of those relationships were heavily influenced by subjects' responses to previously unaddressed FQ topics. Yet, it appears that people who pursue social approval were still monitoring their responses in order to maximize accuracy by filtering out inaccurate details when no metamemory instructions were provided. Conscientious people provided details that were less precise in the IDK+ condition, which is a relationship that was specific to the FQ Mention analyses. Perhaps when conscientious people were given directions regarding report option, it made them more cautious with their metamemory decisions, leading to details that were less precise. This relationship may not have been significant in the FQ All analyses because when FQs broached new topics, people in this condition may have opted to say "I don't know" rather than to adjust the precision of their responses.

Theoretical Implications

There are three theoretical implications from the results of this dissertation: (1) Metamemory instructions may influence people's search for details, not just their accuracy and informativeness criteria. (2) Metamemory instructions, or the lack thereof, may have a different effect on FNs than on FQRs. (3) Each interview format may primarily rely on different types of processing.

A previous interview study that used the same movie clip and coders resulted in FNs that were significantly more accurate than FQRs (Hopkins et al., 2017), whereas this dissertation resulted in FNs with relatively low accuracy, statistically the same as FQR accuracy. One possible explanation for the unexpectedly poor FN accuracy is the

differences in FN instructions between experiments. Subjects were asked to describe what happened "in as much detail as [they could]" in the Hopkins et al. study. These instructions could have affected subjects' decisions about which details to report, so this statement was removed from the FN instructions for this dissertation in order to isolate the effect of interview format more clearly (see Appendix A).

Another statement was removed from the Hopkins et al. (2017) FN instructions because of concerns about influencing subjects' decision-making process. The Hopkins et al. instructions also told subjects that "any piece of information could potentially be important." As discussed in Chapter I, people not only set an accuracy criterion when deciding what to report, they also set an informative criterion. If something is deemed as too vague to be worth mentioned (even if it meets the accuracy criterion), people often choose not to report that detail (Ackerman & Goldsmith, 2008). Therefore, this second statement was removed from FN instructions because it essentially removes the informative criterion from the decision-making process.

It is not always clear what instructions preceded FNs in previous studies, but of the studies that provided enough information to determine what the FN instructions were, all but one study included metamemory instructions (Evans & Fisher, 2011; Taylor & Dando, 2018). While Odinot et al. (2009) seems to be the only one that did not include metamemory instructions prior to the FN, the subjects were real victims of crime who had previously been interviewed by police, and the subjects were provided a floor plan of the scene of the crime to aid their memory. Therefore, it is possible that the results of this dissertation indicate the effect of interview format on detail quality, when influential instructions are not provided before FQs *or FNs*.

This dissertation may be the first interview study that did not include metamemory instructions prior to FNs, and it is the first one to my knowledge that did not result in high-accuracy FNs. The Informative+ FQ instructions most closely align with typical FN instructions (e.g., "tell me everything you can remember"), but while the Informative+ and Control instructions resulted in similar FQR quality, Informative+ (other studies) and Control FN instructions (this dissertation) seemed to result in different FN quality. This may mean that metamemory instructions affect FNs differently than FQRs. If metamemory instructions influence interview formats differently, it is unclear why this happens.

One possible explanation is that while FQ metamemory instructions influence the accuracy and informativeness criterion in the decision-making process, FN metamemory instructions affect an earlier step: searching for details. Assuming FN metamemory instructions do affect a different stage of the decision-making process than FQ metamemory instructions, it also suggests that people are primarily relying on different forms of processing in each interview format. People may primarily use top-down processing during FNs, meaning that any search for details from memory is self-guided, and that encouraging people to expand or prolong their search is helpful. When people answer FQs their search is directed more by others, meaning that more bottom-up processing may be used in their search for details.

Limitations

Any explicit instructions regarding how metamemory should be utilized were removed from the Hopkins et al. (2017) FN instructions for this dissertation, however additions were made to indicate to subjects that they should consider themselves

eyewitnesses. Specifically, subjects were reminded that they witnessed a crime in the video, and they were told that the police were interviewing them to help solve the case. The word *crime* was chosen over the word *con* in order to avoid introducing a topic that would be addressed in an FQ, but it may have unintentionally caused people to focus on a specific aspect of the con, since some of the fifteen steps of the *con* involved a clear *crime*. The con in the movie clip involves 15 steps and the end result is the brothers getting part of the drycleaner's profit because they generated business for him. However, one or more of those 15 steps involved larceny—an aspect of the con that was not covered in the FQs, and thus not coded for in the FNs. If more of the FN details were coded for, different patterns in quality might emerge.

Additionally, some studies have found that when questions are administered on a computer, people who score high on the social desirability scale are less likely to present themselves in a better light than when the same questions are administered face-to-face (e.g., Zhang, Kuchinke, Woud, Velten, & Margraf, 2017). However, other studies have found that how questions are administered does not affect how people with high social desirability scores respond (e.g., Persoskie & Nelson, 2013). Thus, the relationships between social desirability and quality of FQRs found in this dissertation may not be representative of a face-to-face interview.

Finally, there are limitations for applying these findings to real-world situations. The crime witnessed was a movie clip, so aspects of the experience such as emotional reactions were not comparable to a real-world experience. Also, the interview was conducted by typing answers on a computer, rather than giving an oral account to a person. Furthermore, the crime in question was a con that was carried out over several

days, meaning that some central characters' appearance changed, and several locations were involved. While some crimes may fit this description (e.g., white-collar crimes, long-term abuse), many do not (e.g., hit-and-run accident, burglary).

Future Directions

While one of the aims of this dissertation was severely limited by the low-accuracy FNs subjects provided, this odd result also generated some interesting questions to be explored in future studies. First and foremost, future studies should examine whether metamemory instructions affect the quality of details differently depending on the interview format the metamemory instructions precede. Although FN instructions used in previous studies were not always described, it appears that this dissertation may be the first to not include metamemory instructions prior to FNs (e.g., "tell me everything you witnessed"), and also the first to result in low-accuracy FNs. Examining the effect of metamemory instructions on FNs is a necessary first step to not only better understand the results of this dissertation, but also to begin to understand whether and how attentional processing may differ between interview formats.

Once the impact of metamemory instructions on FNs is determined, the FN instructions that result in the highest-quality details should be combined with varying FQ metamemory instructions. Perhaps when FN metamemory instructions (e.g., anything could be important, report everything) conflict with FQ instructions (e.g., try to remain accurate, say "I don't know" if needed), witnesses become confused as to how they should utilize their metamemory when answering FQs. Combinations of metamemory instructions prior to FQs should also be examined. Currently, it is unclear whether IDK+ instructions selectively filter inaccurate FQR details, or reduce overall output. Perhaps if

IDK+ instructions were combined with Informative+ instructions, subjects would report more accurate details and still filter out inaccurate details.

Finally, it would be helpful to see how real witnesses respond to FNs and FQs. Interview procedure studies could be replicated using virtual reality so that subjects could experience being a victim or a witness to a crime (e.g., a valuable item being stolen) to make the experience more realistic, but still in a controlled environment, to determine the external validity of current findings. Another option is to use body camera footage that documents actual on-the-scene interviews of real eyewitnesses by responding officers. When there is footage of a crime being committed (e.g., from a security camera) and body camera footage of an eyewitness interview being conducted at the scene of the crime, detail quality could be coded for using the same variables as this dissertation. While both of these options would be time-consuming and difficult, the results could be highly useful.

Conclusions

This dissertation provided evidence that the quality of details subjects report in FQRs differs from those provided in FNs. However, attempting to explain how this change in quality is responsible for the typical difference in accuracy between formats is impossible, as accuracy was the one variable not affected by interview format. Additionally, subjects in the IDK+ condition provided statistically fewer inaccurate details than subjects in the Control condition but the number of accurate details did not differ between conditions. While police may want to incorporate instructions that explicitly encourage "I don't know" responses prior to FQs in situations where errors could cause significant problems (i.e., no corroborating evidence), they must also

consider the need for a high number of details from witnesses in these same situations.

More research is needed to determine the nuances of the effect of IDK+ instructions.

While there are no clear-cut interview protocol changes that should be put into place as a

result of this dissertation, future studies that examine the implications of this dissertation

could lead to evidence-based changes to improve the quality of details witnesses provide

in interviews.

REFERENCES

Ackerman, R. & Goldsmith, M. (2008). Control over grain size in memory reporting –
with and without satisficing knowledge. *Journal of Experimental Psychology:
Learning, Memory, and Cognition, 34,* 1224-1245, doi: 10.1037/a0012938

Bergman, R., Japhet, W., Stern, J.D. (Producer), & Johnson, R. (Director). (2008). *The
Brothers Bloom* [Motion Picture]. United States: Summit Entertainment. \

Blumstein, A. & Cohen, J. (1979). Estimation of individual crime rates from arrest
records. *Journal of Criminal Law and Criminology, 70,*561-585.

Cady, H. M. (1924). On the Psychology of Testimony. *The American Journal of
Psychology, 35,* 110-112.

Center for the Study and Prevention of Violence (1995). *Lies, Dam Lies and Arrest
Statistics*. Boston, MA: Elliot. D. S.

Crown, D. P. & Marlowe D. (1960). A new scale of social desirability independent of
psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Delaney, H. D., & Maxwell, S. E. (1981). On using analysis of covariance in repeated
measures designs. *Multivariate Behavioral Research, 16*, 105–123.

Dietze, P. M., Sharman, S. J., Powell, M. B. & Thomson, D. M. (2013). Does free recall
moderate the effect of mental context reinstatement instructions on children's
cued recall? *Psychology, Crime & Law, 19,* 881-891.
doi: 10.1080/1068316X.2012.700312

Evans, J. R. & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision, and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology, 25,* 501-508. doi: 10.1002.acp.1722

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. doi: 10.1016/j.jml.2005.01.010

Goldsmith, M., Koriat, A., & Weinberger-Eliezer, A. (2002). Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: General, 131,* 73-95. doi:10.1037//0096-3445.131.1.73

Hilgard, E. R. & Loftus, E. F. (1979). Effective interrogation of the eyewitness. *The International Journal of Clinical and Experimental Hypnosis, 27,* 342-257.

Hopkins, R. F. & Lyle, K. B. (2017). Why are People Less Accurate When Answering Follow-Up Questions Than When Freely Recalling a Criminal Event? Poster to be presented at the fifty-eighth annual meeting of the Psychonomic Society. Vancouver, Canada.

Humphries, J.E., Holliday, R. E., Flowe, H. D. (2012). Faces in Motion: Age-Related Changed in Eyewitness Identification Performance in Simultaneous, Sequential, and Elimination Video Lineups. *Applied Cognitive Psychology, 26*, 149-158 doi: 10.1002/acp.1808

Inabau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2015). *Essentials of the Reid Technique: Criminal Interrogation and Confessions* (2nd ed.). Burlington, MA: Jones & Bartlett Learning.

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,* 490-517.

Koriat, A., Goldsmith, M., Schneider, W., & Nakash-Dura, M. (2001). The credibility of children's testimony: Can children control the accuracy of their memory reports? *Journal of Experimental Child Psychology, 79,* 405-437. doi: :10.1006/jecp.2000.2612

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York: Guilford Press.

Lipton, J. P. (1977). On the Psychology of Eyewitness Testimony. *Journal of Applied Psychology, 62,* 90-95.

Marquis, K. H., Marshall, J., & Oskamp, S. (1972). Testimony validity as a function of question form, atmosphere, and item difficulty. *Journal of Applied Social Psychology, 2,* 167-186.

McBee, M. & Field, S. H. (2017). Confirmatory study design, data analysis, and results that matter. In Make, M. C. & Plucker, J. A. (Eds.), *Toward a More Perfect Psychology: Improving Trust, Accuracy, and Transparency in Research* (pp. 59-78). Washington, DC: American Psychological Association.

Memon, A., Wark, L., Bull, R., & Koehnken, G. (1997). Isolating the effects of the
    cognitive interview techniques. *British Journal of Psychology, 88*, 179-197.

Nettle, D. (2006). The evolution of personality variation in humans and other animals.
    American Psychologist, 61, 622-631. doi: 10.1037/0003-066X.61.6.622

Odinot, G., Wolters, G., & van Koppen, P. J. (2009). Eyewitness memory of a
    supermarket robbery: A case study of accuracy and confidence after 3 months.
    Law and Human Behavior, 33, 506-514. doi: 10.1007/s10979-008-9152-x

Persokie, A. & Nelson, W. L. (2013). Just blowing smoke? Social desirability and
    reporting of intentions to quit smoking. *Nicotine & Tobacco Research, 15,* 2088-
    2093. doi: 10.1093/ntr/ntt101

Scoboria, A. & Fisico, S. (2013). Encouraging and clarifying "Don't know" responses
    enhances interview quality. *Journal of Experimental Psychology: Applied, 19,* 72-
    82. doi:10.1037/a0032067

Sharps, M. J., Herrera, M., Dunn, L., & Alcala, E. (2012). Repetition and reconfiguration:
    Demand-based confabulation in initial eyewitness memory. *Journal of
    Investigating Psychology and Offender Profiling, 9,* 149-160.
    doi: 10.1002/jip.1361

Skogan, W. G. & Antunes, G. E. (1979). Information, apprehension, and deterrence:
    Exploring the limits of police productivity, *Journal of Criminal Justice, 7,* 217-
    241.

Weber, N. & Brewer, N. (2008) Eyewitness Recall: Regulation of grain size and the role
    of confidence. *Journal of Experimental Psychology: Applied, 14,* 50-60.
    doi: 0.1037/1076-898X.14.1.50

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

Yarmey, D. A. & Yarmey (1997). Eyewitness Recall and Duration Estimates in Field Settings. *Journal of Applied Social Psychology, 27*, 330-344.

Zhang, X. Kuchkinke, L., Woud, M. L., Velten, J., Margraf, J. (2017). Survey methd matters: Online/offline questionnaires and face-to-face or telephone interviews differ. *Computers in Human Behaviors, 17*, 172-180. doi: 10.1016/j.chb.2017.02.006

APPENDIX A

Free Narrative Directions:

You have just witnessed a crime. Imagine a police detective is interviewing you to help

them solve the case, please describe what you witnessed.

Follow-up Question Directions (Control):

You will now be asked some questions regarding the crime you witnessed.

Follow-up Question Directions (IDK+):

You will now be asked some questions regarding the crime you witnessed. If you can't

provide any additional information, simply respond "I don't know/remember."

Follow-up Question Directions (IDK-):

You will now be asked some questions regarding the crime you witnessed. Please try to

avoid responding "I don't know/remember" unless absolutely necessary.

Follow-up Question Directions (Accuracy+):

You will now be asked some questions regarding the crime you witnessed. It is

extremely important that any additional information you provide is as accurate as

possible.

Follow-up Question Directions (Informative+):

You will now be asked some questions regarding the crime you witnessed. It is

extremely important that any additional information you provide is as informative (e.g.,

helpful, useful) as possible.

APPENDIX B

Follow-up Questions

1. People often report that the younger boy was attracted to a girl in town. Can you provide any details about this girl?

2. People often report that the boys generated a light in the cave scene. Can you provide any details about the light source?

3. People often report that the boys made a deal with a drycleaner. Can you provide any details about this person?

4. People often report that children had frozen treats in the clip. Can you provide any details about where the treats were bought and who sold them?

5. People often report that the boys came up with a plan to con people out of money and that they wrote their plan on a piece of paper. Can you provide any details on this piece of paper?

6. People often report that the boys were going somewhere at the end of the clip. Can you provide any details about what they were doing?

APPENDIX C

Manipulation Check Questions

1. Please write a short summary of how you were instructed to answer the follow-up

    questions (after you provided your initial statement).

2. Before this experiment, had you previously seen the movie *The Brothers Bloom*?

APPENDIX D

Coding Rubric

GIRL
Relevant segments: 1:34-2:00, 2:32-2:50, 3:20-3:58, 4:09-4:48
FQ: People often report that the younger brother was attracted to a girl in town.
Can you provide any details about this girl?

	Accurate Detail	Correct Category-Inaccurate Detail	Incorrect Category-Inaccurate Detail
Precision 1 (General)	**Clothes:** Nice/dressy/modest/ light-colored (1) **Appearance:** Average height (without age/age range) (1), light-colored eyes (2), light-colored hair (2) **Actions:** playing (1), in park/field (1)		
Precision 2	**Clothes:** Mostly/typically dresses (2), wearing pink (2) **Appearance:** Tall (1), thin (1), young (1), average height (with age/age range) (1), about the same height as younger boy (3) **Actions:** playing in park/field (2)	**Clothes:** always wore dresses (2)	
Precision 3	**Clothes:** Skirts and dresses (2), with designs (1) **Appearance:** Blonde hair (2), hair pulled away from face (2), taller than Bloom (2), white (1), under 100 lbs (2) **Actions:** at park with group of kids (5) **Quotes:** asked about the cave (2), asked about the willowisp	**Appearance:** brown hair (2)	**Appearance:** red hair (2), 3 ½-4 feet (2)

	(2), told someone to shut up (3)		
Precision 4	**Clothes:** short-sleeved dresses (2), past knee length (3) **Appearance:** pinned to the side of her head (3), wavy/curly hair (2), golden hair (2), about Bloom's age/roughly 10 (2), a little taller than Bloom (3), about/just under 5ft (2/3), blue/green eyes (1) **Actions:** talking/standing/dancing/ running in field/park with friends (4), holding/picking a flower (2)	**Appearance:** pigtails/pony tail (1), braids (1)	**Appearance:** brown eyes (2)
Precision 5 (Specific)	**Clothes:** white skirt (2), red/pink & white striped top (4), white/gray/cream dress (2), brown/red/rust pullover (2) pink and white dress (3), white shoes/socks (2) **Appearance:** strawberry/ light/dirty blonde (2), light brown (2), middle part (2), pale (1), no freckles (1), fair-skinned (1) **Actions:** swiping at bubbles (2), with boy and girl (3), talking to girl (2), falling in cave (2), reaching back/waiting for Bloom (2), holding hands with Bloom (2), 1st to go in cave with Bloom (3) **Quotes:** "what kind of cave? (4), "Shut up, Dave" (3), "The willowisp?"(2), "Just like the hermit said" (5)	**Clothes:** dark pink skirt in first scene (4) **Appearance:** **Actions:** talking to boy in field (3), blowing bubbles (2)	**Clothes:** polka dot dress in first scene (4) **Appearance:** bow in hair (2), freckles (1) **Actions:**

LIGHT

Relevant segment: 4:33-4:40

FQ: People often report that the boys generated a light in the cave scene.
Can you provide any details about the light source?

	Accurate Detail	Correct Category-Inaccurate Detail	Incorrect Category-Inaccurate Detail
Precision 1 (General)	**Person:** a person (1), somebody (1) **Source:** hand-held light (2)		
Precision 2	**Person:** a child (1), he (1), someone they/the children knew (2)	**Person:** "other person"-specifically not the brothers (2)	
Precision 3	**Person:** a boy (1) **Source:** like/similar to a flashlight (2)		
Precision 4	**Person:** One of the brothers (2), 13 year old boy (2), the older boy (2) **Source:** a flashlight (1)	**Source:** lantern (1), torch (1), candle (1)	**Source:** a big projector (2)
Precision 5 (Specific)	**Person:** Stephen (1), the older brother (2), Bloom's brother (2) **Source:** Battery-powered flashlight (2), yellow flashlight (2)	**Person:** Bloom (1), the younger brother (2) **Source:** electric lantern (2), standard lighter (2)	**Source:** kerosene lantern (2), the sun (1)

DRYCLEANER (PERSON)

Relevant Segment: 5:40-5:45

FQ: People often report that the brothers made a deal with a drycleaner.
Can you provide any details about this person?

	Accurate Detail	Correct Category-Inaccurate Detail	Incorrect Category-Inaccurate Detail
Precision 1 (General)	Adult (1), dark clothes (2), dull-colored clothes (2)		
Precision 2	Older (1), he/man (1), large (1), wearing blue (1)		
Precision 3	Overweight (1), average height/tall (1), white (1), middle-age (1), medium build (2), not lean (1), dressy pants/slacks (2), dressy/nice shirt (2), "old, but no trouble walking/not bent over" (3), "not long hair, but not a buzz cut" (4)		Facial hair (2), dark hair (2)
Precision 4	Balding/receding hairline (1), tan (1), taller than Steven (2), a little chubby (2), range of age that falls in middle-age (2), round in stomach (2), dark/striped shirt (2), long pants (2), about 5' 9"/around 6" (2)	Bald (1), blonde hair (2), late 40s-early 40s (4)	Mustache (1), beard (1), glasses (1), brown hair (2), had cash in hands (2), dirty clothes (2)
Precision 5 (Specific)	Short/white/salt and pepper/gray hair (2), bald on top/hair on sides (2), black pants/slacks (2), dark blue and dark green (4)/vertical striped (2)/button-up (1) /untucked (1) /collard (1) /short-sleeved (1) shirt (1), folded arms (2)		

FROZEN TREATS (WHERE/WHO)

Relevant Segment: 1:07-1:15

FQ: People often report that children had frozen treats in the clip.

Can you provide any details about where the treats were bought and who sold them?

	Accurate Detail	Correct Category-Inaccurate Detail	Incorrect Category-Inaccurate Detail

Precision 1 (General)	**Where:** Store/shop (1)		**Who:** Person (1)
Precision 2	**Where:** Place where they bought candy (2)		**Who:** He/guy/man (1)
Precision 3	**Where:** Small store (2), local store (2), stand-alone store (2)		**Who:** Older (1), white (1), white outfit (2), apron (1)
Precision 4	**Where:** Candy store (2), ice cream store/parlor (2), about 2 steps (3), yellow sign (2)	**Where:** Snack shop (2)	**Who:** Clerk (1), the owner of the dry cleaners (3)
Precision 5 (Specific)	**Where:** Sign said "Sweets, candy, & Cookies JJs" (5) Read, white, and yellow sign (4), white (1) brick (1) and cement (1), **Who:** You never see a person/it's unclear who they purchased it from (2)	**Where:** Convenient store (2), general store (2), corner store (2), ice cream truck (2), grocery store (2)	**Where:** It never showed where (3), gas station (1), dry cleaners (1), retail store (2), drug store (2) **Who:** O'Henry (1), white button-down shirt (3)

PLAN/CON (STEPS & PAPER)
Relevant Segments: 2:18-2:26, 2:51-3:00, 5:28-5:38
FQ: People often report that the brothers came up with a plan to con people out of money and that they wrote their plan on a piece of paper.
Can you provide any details on this piece of paper?

	Accurate Detail	Correct Category-Inaccurate Detail	Incorrect Category-Inaccurate Detail
Precision 1 (General)	Steps (1), methods (1), instructions (1) guideline (1), Step 6 was important (2) Left paper when they left (4) The rest of the steps Bloom's brother did a lot of things (3)		
Precision 2	Large/medium paper (1), Multiple/multi-step (1), long (1), ordered steps (2), list/list of things to do, detailed plan, Included [generic descriptions of steps] (e.g., made deal with drycleaner)		Normal-sized piece of paper

Precision 3	White (1), many/several steps (1), every step (1), nonlinear (1)	Paid off drycleaner (2)	
	Smaller than poster (3), numbered steps (2), Last seen in foster home (4), Mention what happens for a steps without identifying which # (e.g., one step was for Bloom to talk to the girl)		
Precision 4	Unlined/plain paper (1), black ink (2), algorithm (1), sequence (1), flow chart (1), range of steps (e.g., 13-15) that includes 15 (2)	Web (1), list (1), folded into sixths (2)	Map (1), lined (1), loose leaf (1)
	A bit larger than printer paper (3), play-by-play (1). Diagram (1), step by step (2), boys left it on the floor of thier room (3), started with bloom talking to girl (4), Step 1 to talk to girl (4), Exchange for a cut of revenue (3		
Precision 5 (Specific)	Sketchbook (1), frayed edge (1), circled (1)/ bubbled (1)/ squared (1) steps, line connecting steps (2), 15 steps (2), handwritten/ drawing/wrote (1), marker Filled page (2), all steps on front (3), clear back (2) Horizontal/landscape orientation (2), First seen on stump in woods (5), Last seen crumpled under the bed (4), #1 "Bloom talks to girl" (4), #3 Stephen finds a cave (3), #5 "Stephen buys supplies" (4), #6 "cut % O'Henrys" (3), #8 "Stephen scouts church" (4)	Printer/copy paper (1), construction paper (1), pen (1), pencil (1), lists specific % (2), lists specific number of steps that's not 15 (but close) (2), investing in drycleaners (2)	Notebook paper (1)

LEAVING TOWN

Relevant Segment: 5:45-6:26

People often report that the brothers were going somewhere at the end of the clip.

Can you provide any details about what they were doing?

	Accurate Detail	Correct Category-Inaccurate Detail	Incorrect Category-Inaccurate Detail
Precision 1 (General)	Had suitcases/luggage/bag/baggage (2)		Being transported/picked up (1), They left (2)
Precision 2	Both carrying suitcases/luggage/bag/baggage (3), eating frozen treats (2)	Briefcase/duffle bag (1)	Forced to move/leave (2), live on their own/running away (2), somewhere far away (2), next task (2), send them back (3)
Precision 3	Eating ice cream (2), dark-colored suitcases (2)		Going to their next home (2), going to pull their next con (2), different town (2)
Precision 4	Any of the specific info with just listing "one of the brothers" instead of name, Brown suitcases, left their pops and started walking (4), eating popsicles (2), both wore suits and hats (3)		Leaving the laundry mat (2), leaving the candy/ice cream store (3), going to foster care (2), returning to a previous town (2), going to a boys home (2), put back up for adoptions (2), got into a car (2), heading out of town (2)
Precision 5 (Specific)	Slowly walking (2), in alley/on sidewalk (1), Bloom was sitting on steps (3), Stephen throws rocket pop over shoulder (4), Bloom looks back at children (3), Bloom struggles to carry his bag (3), eating rocket pops (3), Stephen had 1, bloom had 2 (4), it did not state where they were going (3)	Firecracker popsicles (2)	At bus stop/got onto train (2), CMS/social services/adoption center/Juvenile detention center/child services/social worker/ real parents (1)

CURRICULUM VITAE

Robin F. Hopkins
University of Louisville
Department of Psychological and Brain Sciences
Louisville, KY 40292
859-576-0656
robin.hopkins@louisville.edu

EDUCATION

M.S. in Experimental Psychology
University of Louisville, Louisville, KY
2015

B.S. in Psychology
High Point University, High Point, NC
2012

PUBLICATIONS

Published
Lyle, K. B., Dombroski, B. A., Faul, L., **Hopkins, R. F.**, Naaz, F., Switala, A. E., Depue, B. E. (2017). Bimanual Coordination Positively Predicts Episodic Memory: A combined Behavioral and MRI Investigation. *Brain and Cognition, 118*, 71-79. doi: 10.1016/j.bandc.2017.07.009

Hopkins, R. F., Lyle, K. B., Hieb, J. L., Ralston, P. A. S. (2016) Spaced Retrieval Practice Increases College Students' Short- And Long-Term Retention of Mathematics Knowledge. *Educational Psychology Review, 28*, 853-873. doi: 10.1007/s10648-015-9349-8

Edlin, J. M., Leppanan, M. L., **Fain, R. J.**, Hackländer, R. P., Hanaver-Torrez, S.D., & Lyle, K. B. (2015*)*. On the use (and misuse?) of the Edinburgh Handedness Inventory. *Brain and Cognition, 94,* 44-51. doi: 10.1016/j.bandc.2015.01.003

In Progress (Data Collected)
Hopkins, R. F., Bego, C. R., Ralston, P. A. S., Hieb, J. L., & Lyle, K. B. (In Progress). Retrieval Practice and Spacing:

Effects on Long-Term Learning among Engineering
Precalculus Students.

Hopkins, R. F. & Lyle, K. B. (Under Review). Don't Blame the
Bartender: Effects of Image-Size Disparity on Face
Matching. *Applied Cognitive Psychology*

CONFERENCE
PRESENTATIONS
AND INVITED
TALKS

Hopkins, R. F., Lyle, K. B., Ralston, P. A. S., & Hieb, J. L.
(2018). Retrieval practice and spacing: Effects on
precalculus learning among engineering students. Poster to
be presented at the one-hundred twenty-fifth annual
meeting of American Society for Engineering Education,
Salt Lake City, UT

Hopkins, R. F., Pinkston, M., & Lyle, K. B. (2017). Why are
People Less Accurate When Answering Follow-Up
Questions Than When Freely Recalling a Criminal Event?
Poster presented at the fifty-eighth annual meeting of the
Psychonomic Society. Vancouver, Canada.

Hopkins, R. F. & Lyle, K. B. (2017). Getting Away with Fake
IDs?: Effects of Decision Time and Difference Prevalence
on Face-Matching Errors. Poster presented at the twenty-
ninth annual meeting of the Association for Psychological
Science, Boston, MA

Hopkins, R. F. (2016). Why Students Forget Most of What They
are Taught and What You Can Do About It. Academic
Services Tutor Training. High Point University, High
Point, NC.

Lyle, K. B., Depue, B. E., Dombroski, B. A., Switala, A. E.,
Hopkins, R. J., & Leppanen, M. L. (2016). Bimanual
coordination positively predicts episodic memory: A joint
behavioral and neuroimaging investigation. Paper
presented at the fifty-seventh annual meeting of the
Psychonomic Society. Boston, MA.

Fain, R. J. & Lyle, K. B. (2015). Don't blame the bartender:
Image-size disparity increases false face-matches. Poster
presented at the fifty-sixth annual meeting of the
Psychonomic Society. Chicago, IL.

Fain, R. J., Hieb, J. L., Lyle, K.B., & Ralston, P. A. S. (2015).
Can the Spacing Effect Improve the Effectiveness of a
Math Intervention Course for Engineering Students?
Poster presented at the one-hundred twenty-second annual
meeting of American Society for Engineering Education,
Seattle, WA.

Fain, R. J., Lyle, K. B., Hieb, J. L., & Ralston, P. A. S. (2015).
Spaced Retrieval Practice Increases Retention of Calculus
in an Engineering Mathematics Class. Poster presented at

the twenty-seventh annual meeting of the Association for Psychological Science, New York, NY.

Leppanen, M. L., **Fain, R. J.**, Edlin, J. M. Hackländer, R. P., & Lyle, K. B. (2014). How do researchers use the Edinburgh Handedness Inventory? Poster presented at the eighty-sixth annual meeting of the Midwestern Psychological Association, Chicago, IL.

Fain, R. & Wear, K. (2012). Eye Remember: Continuing to Improve Episodic Memory. Poster presented at the meeting of the NC Cognition Group, Chapel Hill, NC.