6-3-2014

An Investigation of Factors that Create and Mitigate Confirmation Bias in Judgments of Handwriting Evidence

Jeffrey Paul Kukucka
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: http://academicworks.cuny.edu/gc_etds
Part of the Psychology Commons

Recommended Citation
http://academicworks.cuny.edu/gc_etds/239

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects (2014-Present) by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@gc.cuny.edu.
AN INVESTIGATION OF FACTORS THAT CREATE AND MITIGATE
CONFIRMATION BIAS IN JUDGMENTS OF HANDWRITING EVIDENCE

by

JEFFREY P. KUKUCKA, JR.

A dissertation submitted to the Graduate Faculty in Psychology
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy, The City University of New York

2014
This manuscript has been read and accepted by the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy

Saul M. Kassin

Date

Chair of Examining Committee

Maureen O’Connor

Date

Executive Officer

Emily Balcetis

Itiel Dror

Maria Hartwig

Maureen O’Connor

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

AN INVESTIGATION OF FACTORS THAT CREATE AND MITIGATE CONFIRMATION BIAS IN JUDGMENTS OF HANDWRITING EVIDENCE

by

JEFFREY P. KUKUCKA, JR.

Advisor: Professor Saul Kassin

Over a century of basic cognitive and social psychological research shows that humans naturally seek out, perceive, and interpret evidence in ways that serve to validate their prevailing beliefs (i.e., confirmation bias; Nickerson, 1998). In criminal justice settings, a priori beliefs regarding the guilt or innocence of a suspect can likewise guide the collection, interpretation, and appraisal of evidence in a self-verifying manner (i.e., forensic confirmation bias; Kassin, Dror, & Kukucka, 2013). Recently, confirmation bias has been implicated as a source of forensic science errors in wrongful conviction cases (e.g., National Academy of Sciences, 2009; Risinger, Saks, Rosenthal, & Thompson, 2002). Accordingly, many have suggested procedural reforms to mitigate the detrimental impact of unconscious bias on judgments of forensic evidence.

Three studies tested the effects of exposure to case information and evidence lineup use on judgments of handwriting evidence in a mock investigation. In Studies 1 and 2, participants who were aware of a suspect’s confession rated non-matching handwriting samples from the suspect and perpetrator as more similar to each other, and were more likely to misjudge them as having been authored by the same individual. The findings of Studies 1 and 2 thus further raise growing concerns over allowing forensic science examiners access to case information that can unwittingly produce confirmation bias and result in erroneous judgments.

In Study 2, the use of a simultaneous evidence lineup increased choosing rates relative to an evidence “showup,” and produced a corresponding decrease in judgment accuracy. In Study
3, sequential evidence lineups dramatically reduced false identifications relative to simultaneous lineups, without causing a significant reduction in correct identifications. By showing parallel effects between forensic evidence lineup identification and eyewitness lineup identification, Studies 2 and 3 suggest the potential value of evidence lineups as a means of protecting against bias and reducing systematic error in judgments of forensic evidence.
Acknowledgements

This research was generously supported by a grant from the National Science Foundation (SES #1323964).

I owe my gratitude to many people whose efforts and contributions made this dissertation possible. First and foremost, I would like to thank my advisor, Saul Kassin, for five years of indispensable mentorship, support, and guidance. I am privileged to have learned as much as I did from you, and I aspire to emulate your superlative example in my own career.

Special thanks are also due to Maria Hartwig and Maureen O’Connor, who have likewise been exceptional mentors to me throughout my graduate training, as well as Emily Balcetis and Itiel Dror, for their valuable insight and feedback that helped to improve this project.

I would like to acknowledge my AP Psychology instructor, Jeanne Blakeslee, who first sparked my interest in psychology, and my undergraduate advisors, Kerri Goodwin and Maggie Bruck, who opened many doors for me along the way and to whom I am forever indebted.

I owe special thanks to Emily Dow, for her love and patience throughout this process and beyond. Thanks also to my outstanding labmates, especially Brian Wallace, Jenny Perillo, Sara Appleby, and Vicky Lawson, for their support and friendship over the past five years.

Last, but certainly not least, I am tremendously grateful to my parents, Jeff and Carol Kukucka, and grandparents, Richard and Lillian Wohlfort. I cannot even begin to recount all that you have done for me here, and I never would have made it to this point without you.
TABLE OF CONTENTS

CHAPTER 1: CONFIRMATION BIAS .................................................................1
  Effects on Visual Perception .................................................................1
  Effects on Social Perception ...............................................................4
  Causes of Confirmation Bias ...............................................................5
  Overcoming Confirmation Bias ..........................................................8

CHAPTER 2: FORENSIC CONFIRMATION BIAS ........................................10
  Effects on Police Investigators ............................................................10
  Effects on Evidence Collection and Evaluation .................................12
  Effects on the Trial Process ...............................................................16
  Theoretical Models of Forensic Confirmation Bias ..............................18

CHAPTER 3: THE FORENSIC SCIENCES ......................................................21
  Recent Criticism of the Forensic Sciences ..........................................22
  Sources of Bias and Error ....................................................................23
    Exposure to Case Information .........................................................24
    Pressure from Investigators ...........................................................32
    Method of Evidence Presentation ....................................................33

CHAPTER 4: THE CURRENT STUDIES .......................................................38

CHAPTER 5: STUDY ONE METHOD ..........................................................40
  Participants and Design .......................................................................40
  Procedure ..........................................................................................41
  Materials ...........................................................................................43
  Hypotheses .........................................................................................47

CHAPTER 6: STUDY ONE RESULTS ..........................................................49
  Effects of Confession, Similarity, and Time ........................................49
  Time 1 Judgments ...............................................................................51
  Time 2 Judgments ...............................................................................53
  Change Scores ...................................................................................54
  Need for Cognition ............................................................................56
  Self-Reported Influence .....................................................................56

CHAPTER 7: STUDY ONE DISCUSSION ....................................................58

CHAPTER 8: STUDY TWO METHOD ..........................................................63
  Pilot Study ..........................................................................................64
  Participants and Design .......................................................................65
  Procedure ..........................................................................................66
  Materials ............................................................................................68
  Hypotheses .........................................................................................72
# Table of Contents

## Chapter 9: Study Two Results

- Choosing Rates ................................................................. 75
- Judgment Accuracy .............................................................. 77
- Judgment Confidence .......................................................... 79
- Accuracy-Confidence Composite ....................................... 80
- Signal Detection Analysis .................................................. 81
- Diagnosticity .................................................................. 83
- Open-Ended Responses .................................................... 84
- Self-Reported Influence .................................................... 86

## Chapter 10: Study Two Discussion ........................................ 88

## Chapter 11: Study Three Method .......................................... 93

- Participants and Design .................................................... 93
- Procedure .................................................................... 94
- Hypotheses .................................................................. 96

## Chapter 12: Study Three Results .......................................... 97

- Choosing Rates ................................................................. 98
- Judgment Accuracy .............................................................. 99
- Accuracy-Confidence Composite ....................................... 101
- Diagnosticity .................................................................. 102
- Judgment Confidence .......................................................... 103
- Self-Reported Influence .................................................... 103

## Chapter 13: Study Three Discussion ..................................... 105

## Chapter 14: General Discussion ............................................. 107

- Exposure to Case Information ........................................... 107
- Forensic Versus Eyewitness Identifications ........................ 114
- Limitations and Future Directions ...................................... 118

## Tables ........................................................................... 122

## Figures ........................................................................... 133

## Appendices ..................................................................... 141

## References ..................................................................... 168
TABLES

Table 1: Frequency of handwriting judgments by presentation, context, and target (Study 2) ... 122
Table 2: Frequency of choosing by presentation, context, and target (Study 2) ................. 123
Table 3: Judgment accuracy by presentation, context, and target (Study 2) ....................... 124
Table 4: Signal detection outcomes with sensitivity ($d'$) and bias ($C$) parameters (Study 2) .... 125
Table 5: Conditional probabilities of correct and false IDs, diagnosticity ($dx$) ratios, and percentages guilty, by presentation and context ................................................................. 126
Table 6: Frequency of handwriting judgments by lineup, context, and target (Study 2) ........ 127
Table 7: Frequency of choosing by lineup, context, and target (Study 2) ............................ 128
Table 8: Judgment accuracy by lineup, context, and target (Study 2) ................................. 129
Table 9: Conditional probabilities of correct and false IDs, diagnosticity ($dx$) ratios, and percentages guilty, by lineup and context ................................................................. 130
Table 10: Comparison of Steblay et al. (2003) meta-analysis and Study 2 results .............. 131
Table 11: Comparison of Steblay et al. (2011) meta-analysis and Study 3 results .............. 132
FIGURES

Figure 1: Effects of confession and similarity on similarity rating (Study 1) .........................133

Figure 2: Confession X source interaction on self-reported influence (Study 1) ......................134

Figure 3: Context X target interaction on judgment accuracy (Study 2) ...............................135

Figure 4: Context X target interaction on accuracy-confidence composite scores (Study 2) ..........136

Figure 5: Context X source interaction on self-reported influence (Study 2) ...........................137

Figure 6: Context X target interaction on number of handwriting featured cited in open-ended
explanation for handwriting judgment (Study 2) .................................................................138

Figure 7: Lineup X target interaction on judgment accuracy (Study 3) .................................139

Figure 8: Lineup X target interaction on accuracy-confidence composite scores (Study 3) .......140
APPENDICES

Appendix A: Informed Consent .................................................................141
Appendix B: Need for Cognition Scale .....................................................142
Appendix C: Demographic Questionnaire ..................................................144
Appendix D: Mock Examiner Instructions ................................................146
Appendix E: Case Summary ....................................................................147
Appendix F: Handwriting Samples (Study 1) ............................................152
Appendix G: Memo from Investigators (Study 1) .......................................153
Appendix H: Comprehension Test ............................................................154
Appendix I: Pre-Lineup Instructions (Studies 2 and 3) ...............................155
Appendix J: Evidence Lineups and Showups ..........................................158
CHAPTER 1: CONFIRMATION BIAS

Confirmation bias refers to people’s natural tendency to seek out and interpret new information in ways that validate their pre-existing beliefs or expectations (Nickerson, 1998). Basic research indicates that sensory perception is not objective, but rather is colored by the subjective qualities of the perceiver (e.g., Bressan & Dal Martello, 2002; Bruner & Potter, 1964). Similarly, a person’s pre-existing beliefs about others can impact interpersonal judgments (e.g., Darley & Gross, 1983; Rosenthal & Jacobson, 1966), and shape behavior toward others in a self-fulfilling manner (e.g., Darley & Fazio, 1980; Snyder & Swann, 1978; Snyder, Tanke, & Berscheid, 1977). Though often driven by motivation (e.g., Dunning & Balcetis, 2013; Kunda, 1990), confirmation bias is an inherent aspect of human cognition that typically operates outside of conscious awareness (Kunda, 1990; Nisbett & Wilson, 1977).

Confirmation bias is pervasive: Its effects have been traced as far back as Pythagoras’ studies of harmonic relationships in the sixth century B.C. (Nickerson, 1998), and references to it can be found in the writings of William Shakespeare and Francis Bacon (Risinger et al., 2002). It can also be problematic: The impact of confirmation bias is seen in “a significant fraction of the disputes, altercations, and misunderstandings that occur among individuals, groups, and nations” throughout human history, including the witch trials of Western Europe and New England, the prolongation of ineffective medical treatments, the formulation of inaccurate medical diagnoses, and adherence to erroneous scientific theories (Nickerson, 1998, p. 175).

Effects on Visual Perception

The notion that the perceiver plays an active role in perception is hardly a new one. In an 1899 article in Popular Science Monthly, psychophysicist Joseph Jastrow described how the process of seeing entails not only the objective input of our sense organs, but also the subjective input of the “mind behind the eye... which guides them in gathering information, and gives value
and order” to sensation (p. 299). Insofar as the latter component is idiosyncratic, Jastrow argued that individuals with different “mental eyes” could come away with very different impressions of the same physical stimulus (p. 301). To illustrate this point, Jastrow noted examples of optical illusions in which the same stimulus can be interpreted in either of two ways (e.g., the duck-rabbit illusion, which can be perceived either as a duck facing to the left or a rabbit facing to the right), causing the observers’ perceptions to involuntarily fluctuate between these two possible interpretations as the mind tries to impose meaning on the stimulus.

The scientific study of this phenomenon began soon after World War II as the “New Look” theorists, led by Jerome Bruner, showed that perception is not solely a function of the stimulus being perceived (i.e., bottom-up processing), but is also shaped by the internal states (e.g., needs, motivations, expectations) of the perceiver (i.e., top-down processing; Dunning & Balcetis, 2013). As Bruner and Postman (1948) explained, the nature of perception cannot be predicted from the physical properties of the stimulus alone, because “what is perceived reflects the predispositions, goals, and strivings of the organism at the moment of perceiving” (p. 203).

In perhaps the earliest such empirical study, Bruner and Goodman (1947) asked children to estimate the size of U.S. coins from memory, and found that children of lower socioeconomic status overestimated the size of the coins to a greater degree than did their more affluent counterparts, especially for coins of higher denomination. The authors argued that the subjective value afforded to the object by the observer changed their perception of its size, with more valuable objects being seen as larger. In a follow-up study, Bruner and Postman (1948) found that students tended to overestimate the size of both a positively-valued symbol (i.e., a dollar sign) and a negatively-valued symbol (i.e., a swastika) relative to a neutral symbol, suggesting that any a priori valuation of a stimulus could influence its perception.
In another classic study, Bruner and Potter (1964) showed participants photographs of common objects (e.g., a dog, a fire hydrant, etc.) that had been blurred to various degrees, and had them watch as the pictures were gradually brought into focus. Despite the fact that all of the photos were ultimately seen with the same objective degree of clarity, participants were less able to correctly identify the objects when their initial blurriness was greater. The authors speculated that participants readily generated hypotheses about the content of the photos upon first seeing them, and then adhered to these often-incorrect hypotheses as little new information emerged to discredit them. Supporting this explanation, participants were quite skilled at identifying the objects when the same photos were shown going out of, rather than into, focus. Thus, it appears that they formed inaccurate beliefs based on the blurry photographs which later “interfered” with their ability to recognize the objects after they became clearer.

The ambiguity of Bruner and Potter’s blurry photographs was conducive to confirmation bias: Because the sensory input from the stimulus was insufficient to allow its recognition, there was greater opportunity for subjectivity to guide perception. Indeed, much of the early research on top-down effects on perception utilized simple ambiguous (also called “reversible”) figures, such as the duck-rabbit figure (Jastrow, 1899) and the wife/mother-in-law figure (Boring, 1930; Leeper, 1935), which can be perceived in either of two ways (for a compendium of such figures, see Fisher, 1968). More recent studies have demonstrated similar effects using more complex stimuli. For example, Bressan and Dal Martello (2002) showed participants photos of adult-child pairs and asked them to rate the facial resemblance between the adult and child. When led to believe that the adult and child were genetically related, participants saw greater resemblance between them, even when they were not in fact related. In short, there now exists a wealth of evidence that an individual’s prevailing beliefs can influence visual perception.
Effects on Social Perception

Confirmation bias can similarly guide how people form impressions of others. In a series of classic studies by Asch (1946, Experiments 6 and 7), participants gave their impression of a hypothetical person who was described as “intelligent, industrious, impulsive, critical, stubborn, [and] envious.” When these six traits were presented in this order, participants formed positive impressions; however, when these same traits were presented in reverse order, they formed negative impressions. In other words, participants were most heavily swayed by whichever traits were encountered first. This finding is consistent with contemporary research indicating that people form impressions of others quickly and automatically -- in some cases, as quickly as 100 milliseconds after seeing a person’s face (Willis & Todorov, 2006).

In these same studies, Asch (1946) found that participants changed their appraisal of the traits in the middle of the list to match the positive or negative impression implied by the first two traits. For example, “critical” is an ambiguous trait that could be taken to mean “analytical” (positive) or “judgmental” (negative), depending on whether it was preceded by other positive or negative traits. Others have since replicated this “change of meaning” effect (e.g., Hamilton & Zanna, 1974; Watkins & Peynircioglu, 1984) by showing that once people form impressions of others, subsequent information is filtered so as to be consistent with this initial impression.

Behavioral confirmation. Once a person forms an impression of another individual, that impression guides his or her behavior, which can result in a self-fulfilling prophecy that further strengthens the impression. This can be problematic in research settings: If an experimenter is not blind to a study’s hypothesis, he or she may unwittingly influence the results so as to produce the anticipated outcome (Rosenthal, 2002). In an early demonstration of this effect, rats that had been arbitrarily labeled as “maze-bright” learned to run mazes more quickly than those labeled as
“maze-dull,” presumably because these labels created expectations that differentially influenced experimenters’ training protocols (Rosenthal & Fode, 1963). Rosenthal and Jacobson (1966) then extended this effect to humans, as they led elementary school teachers to believe that certain students (who had been selected at random) would intellectually “bloom” over the next year. Consistent with the teacher’s expectations, these randomly-chosen students showed the most improvement on an intelligence test given eight months later.

Similar processes affect informal social interactions as well. For example, Snyder and Swann (1978) had students interview a peer who they were led to believe was either extraverted or introverted. First, when given a list of possible interview questions, interviewers chose more extravert-oriented questions (e.g., “What would you do if you wanted to liven things up at a party?”) when they expected the interviewee to be extraverted, and vice versa. In other words, interviewers developed questioning strategies designed to confirm their expectation of the interviewee. Later, a third group listened to the interviews and judged interviewees as more extraverted when they were questioned by an interviewer who expected them to be extraverted. Thus, interviewers elicited behavior from interviewees that confirmed their expectation.

**Causes of Confirmation Bias**

**Selective hypothesis testing and heuristics.** Some researchers have argued that confirmation bias is the byproduct of a general positive test strategy. Klayman and Ha (1987) proposed that humans naturally test hypotheses by looking for instances in which a given property or event *does* occur rather than instances in which it *does not* occur. Although positive test strategies often lead to efficient and correct judgments, they also produce predictable errors.

This point was illustrated in a classic study by Peter Wason (1960), an English psychologist who first coined the term “confirmation bias.” Wason gave participants a set of
three numbers (2, 4, and 6) that fit an unknown rule (i.e., any three ascending numbers), and tested their ability to discover the rule by generating their own sets of three numbers (i.e., “triples”) and receiving feedback as to whether or not each fit the rule. Participants were much more likely to generate triples that fit with their working hypothesis about the rule rather than ones that did not. As a result of this preference for confirmatory feedback, participants often became highly confident that they had ascertained the correct rule, when in fact they had not.

Just as Klayman and Ha (1987) argued that a positive test strategy is largely useful but can be problematic, the same is true for heuristics -- i.e., shortcuts in decision-making based on simple general rules. Classic studies by Tversky and Kahneman (1974) show how heuristics can lead to error. For example, the anchoring and adjustment heuristic entails making an initial judgment (i.e., an “anchor”) which is then adjusted to arrive at a final judgment. Because people tend not to stray far from the anchor, inaccurate anchors tend to result in error. That is to say, if one’s initial expectation is flawed, their ultimate judgment is apt to be similarly flawed.

Motivated perception. Although confirmation bias is an intrinsic feature of cognition, it can also be driven by people’s goals and desires. Kunda (1990) distinguished between two types of motivational goals: accuracy goals, where the goal is to produce an accurate conclusion, and directional goals, where the goal is to arrive at a particular desired conclusion. In the latter case, people are able to maintain an “illusion of objectivity” which prevents them from recognizing that their cognition has been biased in favor of their directional goal (Kunda, 1990, p. 483).

There is growing support for the notion that directional goals can unconsciously skew perception. Balcetis and Dunning (2006) found that individuals were more likely to perceive whichever of two interpretations of an ambiguous figure would lead to a desirable (rather than aversive) outcome, and that this effect was not due to selective reporting but rather to a genuine
effect of motivation on visual perception. Directional goals can also impact spatial perception, as people tend to judge desirable objects as being nearer than they truly are, and vice versa (Balcetis & Dunning, 2010). This effect likely serves an adaptive function: When desirable objects appear closer, the perceiver becomes mobilized to obtain these objects. Finally, directional goals can likewise impact social perception: People are generally biased toward seeing close relationship partners in a positive light (see Gagne & Lydon, 2004), which again appears to serve an adaptive function (Murray & Holmes, 1997; Murray, Holmes, & Griffin, 1996).

It is important to note that perception is not limitlessly malleable, even when motivation is high. To that effect, Kunda (1990) cautioned that “people do not seem to be at liberty to conclude whatever they want to conclude merely because they want to” (p. 482). Instead, there are reality constraints on perception, such that the evidence in favor of one’s judgment must be sufficient to justify that judgment. Even a strongly desired outcome cannot be rationalized in the face of irrefutable evidence to the contrary (Kunda, 1990). In contrast, ambiguous stimuli are highly conducive to confirmation bias, insofar as they allow one to justify any desired conclusion that is even remotely viable (e.g., Darley & Gross, 1983).

**A two-stage model of confirmation bias.** Darley and Gross (1983) posited a two-stage model to explain the process of confirmation bias, irrespective of motivation. First, a person forms an expectation which acts as a tentative hypothesis. Then, he or she tests this hypothesis against the available evidence in a biased manner that confirms his or her initial expectation. The implication is that expectancy confirmation is not an automatic process, but rather an active one in which the perceiver utilizes bottom-up input to validate their top-down beliefs. If the bottom-up evidence is insufficient to justify those beliefs, confirmation bias does not occur.
To test this theory, Darley and Gross (1983) had participants judge the academic ability of a fourth-grade girl named Hannah. First, participants were led to expect that Hannah’s ability would be either low or high; then, some watched a video of Hannah taking a test in which her true ability was ambiguous. Those who did not see the video rated Hannah’s ability as consistent with a fourth-grade level regardless of their expectations, as they did not have enough bottom-up information to rate it otherwise. Those who did see the video rated her ability in line with their expectations, as either above or below a fourth-grade level -- despite the fact that they viewed the same video. Thus, bias manifests itself only when there is some evidence to justify the expected and/or desired conclusion, and even if that evidence is weak or ambiguous.

**Overcoming Confirmation Bias**

In the face of overwhelming evidence that one’s beliefs can unwittingly guide perception in a self-verifying manner, it logically follows to ask whether confirmation bias can be avoided. Wilson and Brekke (1994) outlined four conditions that must be met for a person to overcome “mental contamination” (p. 117). First, the person must be aware of the unwanted processing. This is a considerable hurdle, as research consistently shows that people have little insight into their own biased thinking. Rather, it appears that people often fail to notice or appreciate the stimuli that influence their cognition and behavior (Nisbett & Wilson, 1977) and are unable to recognize the same biases in themselves that they readily detect in others (Pronin, 2007).

Second, the person must be motivated to overcome biased processing, such that its outcome is seen as undesirable enough to warrant the effort to correct it. Third, the person must be aware of both the direction and the magnitude of the bias, so that any effort to correct it can be properly calibrated. Fourth, the person must have sufficient control over their own cognition and behavior to allow for these to be adjusted in order to correct for the bias.
The model thus presents an admittedly “pessimistic” view, as failure at any one of these four stages means a failed attempt at overcoming bias (Wilson & Brekke, 1994, p. 120). The model has also received empirical support -- most notably from Wilson, Houston, Etling, and Brekke (1996), who found support for the four requisite criteria and illustrated the difficulty of ensuring that all four criteria are met. In sum, although effective methods of remedying bias are theoretically well-understood, they appear difficult to implement in practice.
CHAPTER 2: FORENSIC CONFIRMATION BIAS

Forty years ago, Tversky and Kahneman (1974) reasoned that confirmation bias could operate within the criminal justice system and speculated that “beliefs concerning the likelihood of... the guilt of a defendant” could impact decision-making in forensic and legal settings (p. 1124). Moreover, they posited that the effects of confirmation bias would not be limited to “laymen,” but could also impact the judgments of “experienced researchers” (p. 1130).

As it turns out, these statements were quite prescient. Kassin, Dror, and Kukucka (2013) recently argued for the existence of a forensic confirmation bias, and reviewed burgeoning research to show that pre-existing beliefs and motives can influence the collection and evaluation of evidence during the course of a criminal case. In short, it now appears that confirmation bias can pervade the processes of criminal investigation and adjudication. Belief in a suspect’s guilt can impact the perception, judgment, and/or behavior of police investigators (Hill, Memon, & McGeorge, 2008; Kassin, Goldstein, & Savitsky, 2003; Narchet, Meissner, & Russano, 2011). Such beliefs can likewise guide the collection and interpretation of evidence from eyewitnesses (Hasel & Kassin, 2009), alibi witnesses (Marion, Kukucka, Collins, Kassin, & Burke, 2014), and other outside sources (e.g., Elaad, Ginton, & Ben-Shakhar, 1994; Lange, Thomas, Dana, & Dawes, 2011). Finally, confirmation bias can influence the trial process itself, through its effects on judicial instruction (Halverson, Hallahan, Hart, & Rosenthal, 1997), expert witness testimony (e.g., Murrie, Bocaccini, Guarnera, & Rufino, 2013) and juror decision-making (e.g., Carlson & Russo, 2001; Charman, Gregory, & Carlucci, 2009; Kukucka & Kassin, in press).

**Effects on Police Interrogators**

Once a suspect has been identified, investigators assess the suspect’s veracity, which will determine whether he or she is detained for interrogation (Inbau, Reid, Buckley, & Jayne, 2001;
Kassin, 2005). There is some evidence to suggest that these judgments could be subject to confirmation bias. Johnson, Bush, and Mitchell (1998) had participants rate the believability of a set of autobiographical statements. When told that the statements came from an earlier study, participants rated more detailed statements as more credible. However, when told that they came from a police interview, participants rated more detailed statements as less credible, presumably because they expected the suspect to be lying. Innocent suspects who give detailed statements may thus paradoxically be at greater risk of being misidentified as deceptive and detained for interrogation if the investigator believes a priori that the suspect will try to deceive them (Kassin, 2005). This is particularly worrisome in light of the fact that innocent suspects more often waive their right to silence (Kassin & Norwick, 2004) and are more forthcoming with information (e.g., Hartwig, Granhag, Stromwall, & Kronkvist, 2006) than guilty suspects.

Next, a process of behavioral confirmation is likely to unfold in the interrogation room. Using a paradigm similar to that of Snyder and Swann (1978), Kassin et al. (2003) found that mock interrogators who expected a suspect to be guilty asked more guilt-presumptive questions than those who expected the suspect to be innocent, and were more likely to judge the suspect as guilty regardless of the suspect’s actual guilt or innocence. Later, a third group of participants listened to audiotapes of these interrogations and rated suspects as more defensive when they were interviewed by guilty-expectancy interrogators, again irrespective of the suspect’s actual guilt. Thus, an interrogator who mistakenly believes an innocent suspect to be guilty is likely to elicit behavior from the suspect that strengthens this erroneous belief.

Others have since replicated and extended these findings. Hill et al. (2008) found that innocent suspects were seen as more nervous, more defensive, and less believable when asked guilt-presumptive rather than neutral questions, and were consequently more likely to be judged
as guilty. Narchet et al. (2011) found that interrogators who believed a suspect to be guilty used more *minimization* tactics -- which provide moral justification for the offense, downplay its seriousness, and imply leniency -- as well as more *maximization* tactics -- which exaggerate the seriousness of the offense and the strength of the evidence against the suspect (Kassin & McNall, 1991). In turn, these tactics led suspects to report feeling greater pressure, which predicted the likelihood of an innocent person giving a false confession. It thus appears that an interrogator’s expectation of guilt can create a self-fulfilling prophecy by eliciting behavior from innocent suspects that is seen as indicative of guilt and raising the risk of false confession (Kassin, 2005).

**Effects on Evidence Collection and Evaluation**

A prevailing belief in a suspect’s guilt can also affect how investigators gather evidence and assess its probative value. O’Brien (2009) had mock investigators review a case file of evidence in a criminal investigation; some were asked to stop halfway through to articulate their belief as to whether the suspect was guilty and why. As a result of stating this belief, participants became more likely to pursue lines of investigation that specifically targeted this suspect at the expense of other possible leads, and more likely to recall ambiguous evidence as being indicative of guilt. These findings were replicated by Rassin, Eerland, and Kuijpers (2010) in a sample of law students: Those who formed a preliminary belief in a suspect’s guilt after reviewing a case file were more likely to seek out additional evidence of guilt, whereas those who believed that he was innocent tended to look for information to verify his innocence.

Confirmation bias may direct both the search for evidence and its interpretation. Lit, Schweitzer, and Oberbauer (2011) had teams of scent detection dogs and their handlers search an area for a target scent (i.e., drugs or explosives) that was not truly present, while manipulating both the handler’s expectations as to whether or not the target scent was present, as well as the
presence of a decoy scent (i.e., beef jerky and a tennis ball). Overall, 85% of these searches produced at least one false alarm, which could be explained in either of two ways. First, the dogs may have detected subtle cues that betrayed the handlers’ expectations, leading the dogs to emit an alerting response in the absence of the target scent. Second, handlers’ expectations may have led them to interpret the dogs’ behavior as indicative of an alerting response when it was not.

Charman et al. (2009) provided further support for confirmation bias effects on evidence evaluation. As part of a mock crime investigation, students were shown a computerized facial composite of the perpetrator and asked to decide which of four suspects was most similar to the composite. When told that one of these four suspects (selected at random) had been identified by two eyewitnesses as the perpetrator, student-investigators rated the composite as most similar to whichever suspect had allegedly been identified. Thus, the eyewitness evidence likely created an expectation of guilt that influenced their appraisal of the facial composite evidence.

Eyewitness identifications appear to be similarly contingent on belief in a suspect’s guilt. Hasel and Kassin (2009) had participants witness a mock theft and then try to identify the thief from a target-absent lineup (i.e., a lineup in which the actual thief’s photo was not present). Two days later, participants were given new information about the crime and an opportunity to change their lineup decision. When told that the suspect they identified had since confessed, they grew more confident in their identification. When told that a different suspect had confessed, 61% changed their previous identification -- all of whom now identified the confessor as the thief. Finally, among those who at first correctly reported that the thief was not in the lineup, nearly half identified one of the innocent lineup members as guilty after being told that someone had confessed, even if they did not know which suspect had confessed.
In addition to generating false inculpatory evidence, confirmation bias may result in the suppression of exculpatory evidence that would otherwise demonstrate his or her innocence. For example, alibi witnesses may become reluctant to testify on a suspect’s behalf if made aware of other evidence that suggests the suspect’s guilt. In 1986, John Kogut was a suspect to a rape and murder that was committed on the same night as a party for his girlfriend; as a result, Kogut had several alibis who could testify that he was at the party and was therefore innocent. However, after police informed these alibis of Kogut’s confession (which was given during an 18-hour interrogation and later shown to be false), they began to doubt their recollection of the evening and ultimately all withdrew their support for Kogut (Herbert, 2009). Inspired by this case, Marion et al. (2014) recently conducted a laboratory experiment in which they found that participants who were informed of a suspect’s confession were more likely to recant their corroboration of a true alibi relative to those who were told that the suspect denied guilt.

Even without information to suggest guilt, the mere context of a criminal case may create biasing expectations. Lange et al. (2011) speculated that transcriptions of degraded audiotapes, which are often used at trial, could be colored by the top-down expectations of the transcriber. Supporting this hypothesis, they found that participants who were asked to transcribe mildly-degraded audiotapes made more incriminating transcription errors when told that the tape was taken from a criminal interview as opposed to a job interview (e.g., “I raped her after the party” versus “I ripped her after the party”). Thus, the transcribers’ top-down expectations changed which words they “heard” in the same audiotapes. Notably, this effect disappeared when the tapes were heavily degraded, presumably due to insufficient bottom-up input.

Although Lange et al. did not use experiences transcribers in their study, other studies suggest that training and experience does not render people immune to confirmation bias. Elaad
et al. (1994) posited that confirmation bias can impact polygraph examinations in two ways -- by impacting the manner in which polygraph examiners question interviewees and/or the manner in which they score polygraph charts. To test the latter, they asked ten polygraph examiners to score charts that had been deemed inconclusive by independent raters. For half, they were told that the interviewee confessed to the crime (implying guilt); for the others, they were told that someone else confessed (implying innocence). This information clearly impacted how examiners scored the charts: When led to expect guilt, examiners more often scored the charts as deceptive (9%) rather than truthful (4%); when led to expect innocence, they were often scored as truthful (21%) and never as deceptive. However, in a follow-up study in which examiners scored conclusive charts, this biasing information had no effect on their scoring.

Evidence elasticity. This pattern is consistent with confirmation bias theory, which asserts that ambiguous bottom-up input is subject to distortion whereas clear and irrefutable input is not (Kunda, 1990). Similarly, research on the elasticity of evidence (e.g., Ask, Rebelius, & Granhag, 2008) provides something of a forensic analogue to the basic literature on reality constraints. The term “elasticity” is used to describe the degree to which a particular item of criminal evidence lends itself to a malleable interpretation.

To illustrate this phenomenon, Ask et al. (2008) gave police trainees a case file with evidence that implied the suspect’s guilt. Later, they were given one new piece of evidence (DNA, a surveillance photo, or an eyewitness identification) that either affirmed or refuted their expectation of guilt. Overall, trainees rated this new evidence as more reliable when it affirmed rather than refuted their expectation. However, this was moderated by evidence type: DNA was seen as reliable regardless of whether it implied guilt or innocence, whereas photographic and eyewitness evidence was seen as more reliable when it implied guilt. Despite this, trainees
maintained an “illusion of objectivity” (Kunda, 1990) which was evident in their open-ended explanations of why they believed or disbelieved the new evidence. For example, 18% of those given inculpatory eyewitness evidence noted that the lighting conditions were good, while 24% of those given exculpatory eyewitness evidence noted that the lighting conditions were poor.

**Effects on the Trial Process**

Confirmation bias may also operate during the trial itself, through its effects on judicial instructions, expert testimony, and juror decision-making. With respect to juries, Carlson and Russo (2001) noted that jurors process evidence in a “step-by-step” fashion to arrive at a verdict, and that this process may be subject to predecisional distortion (p. 91). That is to say, jurors who formulate a tentative verdict before they have seen all of the evidence may be more inclined to interpret later evidence as supportive of this verdict. Indeed, early research found that jurors are more heavily swayed by evidence presented earlier at trial (Lawson, 1968), and those who form verdicts early on tend to adhere to these when giving a final verdict (Stone, 1969).

More recent work has explored how jurors’ tentative belief in the defendant’s guilt guide their perception of evidence encountered subsequent to these beliefs. Charman et al. (2009) had mock jurors read a trial summary in which they varied the strength of the evidence against the defendant, and later showed them a facial composite of the perpetrator. As predicted, their self-reported belief in the defendant’s guilt predicted the degree of similarity that they perceived between the defendant and the composite. Thus, jurors who already believed that the defendant was guilty were likely to also see the facial composite as incriminating, whereas those who believed he was innocent saw the facial composite as further evidence of his innocence.

Kukucka and Kassin (in press) tested whether knowledge of a prior confession would taint jurors’ perceptions of handwriting evidence. Mock jurors read a summary of a bank robbery
in which the defendant either denied guilt, or confessed but retracted his confession. They then compared two handwriting samples -- one each from the perpetrator and defendant. When told that the defendant had confessed, jurors saw greater similarity between the handwriting samples, were more likely to incorrectly believe that they were written by the same person, and were more likely to judge the defendant as guilty. The authors then replicated these findings in a repeated-measures design, which showed that mock jurors adjusted their own prior evaluations of the same handwriting stimuli after being informed of the suspect’s confession.

Individual differences may render some jurors more or less susceptible to distorting evidence to match their beliefs. Kassin, Reddy, and Tulloch (1990) found that jurors who were high in need for cognition (NFC) – i.e., the extent to which they enjoy engaging in effortful cognitive activities (Cacioppo & Petty, 1982) – were more persuaded by arguments made early at trial, whereas those low in NFC were more persuaded by arguments presented later at trial. To explain this pattern, Kassin et al. speculated that high-NFC jurors were “overactive” information processors, who were more likely to form an early opinion as to the defendant’s guilt and subsequently find support for this opinion in other ambiguous evidence.

The evidence shown to jurors may be subject to bias and distortion even before it enters the courtroom. For example, expert witnesses with “scientific, technical, or other specialized knowledge” are often hired by litigants to “help the trier of fact to understand the evidence or to determine a fact in issue” (Fed.R.Evid 702). One estimate found that experts testified in 86% of civil jury trials, and came from a variety of fields including medical doctors, psychologists, engineers, and businessmen (Gross, 1991). Given that these experts are hired and compensated by either litigant, they may be inclined to produce testimony that supports that litigant’s case.
This is not a new concern; in tracing the origins of expert testimony as far back as the late 14th century, Learned Hand (1901) expressed concern over a “natural bias” among experts who are called in as the “hired champion... to represent one side and liberally paid to defend it” (p. 53).

Empirical evidence now supports this claim. Gitlin, Cook, Linton, and Garrett-Mayer (2004) collected radiographs from 492 asbestos exposure cases in which the plaintiff had hired an expert radiographer to testify on their behalf, and presented these to a panel of independent radiographers. Whereas hired experts found evidence of abnormalities in 95.9% of cases, the independent experts found such evidence in only 4.5% of the same radiographs. Similarly, Murrie, Boccaccini, Guarnera, and Rufino (2013) demonstrated an “allegiance effect” among psychiatrist expert witnesses, such that those who analyzed sex offender case files on behalf of the prosecution generated higher scores on a measure of risk assessment than those who analyzed the same files on behalf of the defense. Thus, expert witnesses may be biased toward conclusions that support the claim of their litigant. This bias has led some legal scholars (e.g., Robertson, 2010) to propose that expert witnesses be kept blind to which side has retained them.

Lastly, confirmation bias may affect how judges instruct juries, and thereby elicit verdicts that conform to the judge’s expectations. Halverson et al. (1997) had mock jurors instructed by a judge who was led to believe that the defendant was either guilty or innocent, and found that jurors in the former condition were more likely to find the defendant guilty -- even though the evidence presented at trial and the judicial instructions were identical in content.

**Theoretical Models of Forensic Confirmation Bias**

**Tunnel vision.** Martin (2004) described *tunnel vision* as a driving force behind wrongful convictions. Once a suspect is identified, investigators goal is to build a case against that suspect; in so doing, they favor evidence that will help to convict the suspect while “suppressing or
ignoring” evidence that points away from their guilt. Notably, Martin characterizes tunnel vision as an intentional process that stems from pressure on investigators to resolve cases quickly.

Findley and Scott (2006) expanded Martin’s work on tunnel vision by drawing a parallel between tunnel vision and confirmation bias. As such, they claimed that tunnel vision is better understood as a “natural human tendency,” not a product of “maliciousness or indifference” (p. 292). That is to say, investigators unwittingly enter into a confirmatory “feedback loop,” such that their belief in a suspect’s guilt leads them to seek out and favor additional evidence of guilt, which in turn strengthens their belief. Moreover, Findley and Scott posited that tunnel vision impacts not only investigators, but also prosecutors, defense attorneys, and forensic scientists.

In contrast, Snook and Cullen (2008) argued that tunnel vision is generally useful but can lead to error due to constraints on information-processing. They contended that it is unreasonable to expect investigators to pursue every possible suspect and to gather every relevant piece of evidence. Doing so would make investigations less efficient and could decrease the ability to convict the guilty. Notwithstanding the debate over its value, it is clear that tunnel vision can and does shape the course of criminal investigations, as investigators identify a primary suspect and then adjust their information-gathering strategies accordingly.

**Cognitive coherence.** The *cognitive coherence* framework (Holyoak & Simon, 1999) describes confirmation bias in forensic and legal settings as a dynamic interplay between beliefs and evidentiary judgments. Simon (2004; 2011) posited that decision-making entails movement from complexity (i.e., conflicting evidence that must be reconciled) to coherence (i.e., a binary guilt judgment supported by a bulk of the evidence). Coherence is achieved via a bidirectional process in which evidence that is seen as probative points toward a conclusion, while this emerging conclusion concurrently shapes the evaluation of ambiguous evidence such that it
coheres with -- and therein strengthens -- one’s confidence in the emerging conclusion. After many iterations of this process, all of the evidence at hand appears to be internally consistent and unanimously pointing toward the same conclusion (Simon, 2011). To some extent, the cognitive coherence model can be seen as an amalgamation of tunnel vision (Findley & Scott, 2006) -- insofar as one’s beliefs unwittingly guide the valuation of other evidence -- and the Story Model of juror decision-making (Pennington & Hastie, 1986; 1992), which posits that jurors arrive at verdicts by using evidence to construct competing story lines, and choosing whichever story they consider to be most plausible (i.e., comprehensive and internally consistent).

Interdependence among evidentiary judgments produces what Kassin (2012) called *corroboration inflation*: When evidence that implies guilt causes otherwise-neutral evidence to be seen as incriminating, the initial evidence has effectively manufactured additional inculpatory evidence, thus giving the illusion that the strength of the evidence is greater than it truly is. As Simon (2012) put it, “evidence that might otherwise have given rise to a reasonable doubt can be reduced in the fact finder’s mind to a mere negligible doubt” (p. 175). Consequently, when knowledge of one item of evidence corrupts judgments of another, the body of evidence as a whole can become greater than the sum of its parts (see also Charman, 2013).
CHAPTER 3: THE FORENSIC SCIENCES

The collection and appraisal of forensic science evidence is often an integral part of a criminal investigation. The “forensic sciences” encompass many disciplines, including but not limited to firearm and toolmark identification, questioned document examination (i.e., analysis of signatures and handwriting), analysis of trace evidence (e.g., hair, textile fibers, paint chips, etc.), analysis of impression evidence (e.g., fingerprints, shoe prints, bite marks, etc.), blood pattern analysis, arson analysis (i.e., testing for the presence of ignitable materials), and serology (i.e., testing for the presence of bodily fluids; for an exhaustive list of forensic disciplines, see National Institute of Justice, 2006). Many forensic disciplines share a common goal, namely to determine whether a given piece of physical evidence (i.e., a pattern or impression) originated from a particular source. Many of these domains also have a long history, having been used in criminal investigations and presented as evidence in court for roughly a century, where they have been accepted as trustworthy (Mnookin et al., 2011) and seen as highly credible and persuasive by jurors (e.g., Lawson & O’Connor, 2012; Lieberman, Carrell, Miethe, & Krauss, 2008).

However, recent archival analyses have found that forensic science errors are prevalent in known cases of wrongful conviction (e.g., Garrett & Neufeld, 2009). Kassin et al. (2013) argued that confirmation bias may be responsible for many such errors. Given that forensic analyses rely on the visual judgments of human examiners (Dror & Cole, 2010) and that standard practices are likely to cultivate a priori beliefs regarding a suspect’s guilt (e.g., Risinger et al., 2002), these latter beliefs could unintentionally guide the former analyses in a self-verifying manner. Indeed, a growing number of empirical studies show that judgments of forensic evidence are sensitive to examiners’ knowledge of other evidence (e.g., Dror & Charlton, 2006) and other contextual factors (e.g., Dror & Hampikian, 2011; Dror, Peron, Hind, & Charlton, 2005; Miller, 1987).
Recent Criticism of the Forensic Sciences

The forensic sciences have had a predominantly positive impact on the determination of guilt and innocence throughout their history (Risinger, 2010). Yet, several recent high-profile cases involving forensic science errors have led to increased criticism of these disciplines. One notable example was the case of Brandon Mayfield, an American Muslim attorney who was misidentified by multiple independent FBI fingerprint experts in 2004 as the perpetrator of the Madrid train bombings on the basis of latent fingerprint evidence (see Kassin et al., 2013). In 2006, the National Academy of Sciences (NAS) appointed a committee of judges, legal scholars, scientists, and forensic practitioners to scrutinize the current state of the forensic sciences with an eye toward improving standard practices and thereby reducing the incidence of error.

In their ensuing report, the NAS (2009) bemoaned the fact that “the interpretation of forensic evidence is not always based on scientific studies to determine its validity” (p. 8), thus echoing others who have noted the inadequacy of research to support common forensic science practices (Mnookin et al., 2011; Risinger & Saks, 2003). The NAS added that many disciplines lack a clear standardized methodology, which can introduce an element of subjectivity into the analysis that decreases the reliability of judgments and increases the risk of error (see also Dror et al., 2011). In a sweeping dismissal, the NAS concluded that “no forensic method [aside from DNA analysis] has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source” (p. 7). Others have similarly questioned this “faith-based” claim of individualization (Saks & Koehler, 2005, p. 15; see also Saks, 2010), noting that the same source can produce evidence traces that look quite different (e.g., Dror & Cole, 2010), while traces from different sources are often indistinguishable from each other (e.g., Blackwell et al., 2007).
**Contribution to wrongful convictions.** Taken together, these criticisms depict forensic science judgments as tenuous and probabilistic rather than definitively probative. In light of this, perhaps it is not surprising that forensic science errors have now been uncovered in a wealth of wrongful conviction cases. Indeed, such cases were an impetus behind the NAS report, which explained that “faulty forensic science analyses” increase the “risk that true offenders continue to commit crimes while innocent persons inappropriately serve time” (2009, p. 5).

Garrett and Neufeld (2009) analyzed trial transcripts from 137 DNA exoneration cases in which forensic analysts testified at trial, and found that 60% featured forensic expert testimony that either misrepresented or was unsupported by empirical data. In a re-analysis of these data, Hampikian, West, and Akselrod (2010) found that invalid testimony came from a variety of forensic domains, including serology, hair microscopy, bite mark analysis, fingerprint analysis, and DNA analysis. In each of these cases, forensic analysts overstated the probative value of the evidence, and/or failed to disclose exculpatory evidence. In turn, their statements played a role in the convictions of dozens of innocent people (see also Giannelli, 2007).

**Sources of Bias and Error**

The alarming incidence of forensic science error in wrongful conviction cases has stimulated efforts to identify and minimize causes of error. Some errors are likely attributable to examiners who are under-trained or careless, or worse, who knowingly provide false information to investigators or to the courts (e.g., Giannelli, 2007). Though these “bad apples” surely do exist, Thompson (2009) argues that the failures of forensic science are better understood as a systemic problem, rather than a product of incompetence and corruption at the individual level. Similarly, Dror, Kassin, and Kukucka (2013) noted that unconscious biases can impact perception and decision-making even among examiners who are highly-trained, experienced,
competent, and well-intentioned. In fact, errors that stem from misconduct are likely easier to
detect and rectify than those that result from contextual and psychological influences, which
cannot be eliminated by sheer willpower (Dror et al., 2013).

**Exposure to case information.** Risinger et al. (2002) explained that when forensic
evidence is submitted to a laboratory, it is typically accompanied by a transmittal letter that
summarizes the investigation, including other evidence that has been collected. The value of
providing such information to examiners has been the subject of recent debate (see Thompson,
2011). Many forensic examiners have maintained that access to case information is “invaluable”
(Budowle et al., 2009, p. 803) because it helps to guide their forensic analysis and enhances the
accuracy of their judgments (Elaad, 2013), while also “provid[ing] some personal satisfaction
which allows them to enjoy their jobs” (Butt, 2013, p. 60).

In contrast, opponents of this practice have argued that exposure to other evidence that
suggests the suspect’s guilt or innocence can prompt “expectation-laden observations” that
produce “observational errors” (Nordby, 1992, p. 1121). Instead, the proper role of an examiner
is to give a circumscribed judgment of a single piece of evidence that provides independent proof
of the suspect’s guilt or innocence (e.g., Dror et al., 2013; Page, Taylor, & Blenkin, 2012). When
examiners produce accurate judgments that are contingent on their knowledge of other probative
evidence, their accuracy cannot be attributed to any special expertise (Risinger, 2009).

The notion that forensic examiners may be biased by their expectations is hardly a new
one. In an 1894 treatise on distinguishing authentic signatures from forgeries, William Hagan
cautioned examiners as to how confirmation bias can undermine objectivity:

“There must be no hypothesis at the commencement, and the examiner must depend
wholly on what is seen, leaving out of consideration all suggestions or hints from
interested parties... Where the expert has no knowledge of the moral evidence or aspects of the case... there is nothing to mislead him... It is better that the [formulated opinion] be based entirely on what the writing itself shows, and nothing else.” (p. 82)

With this statement, Hagan was perhaps the first to posit that a forensic expert’s preconceptions and motivations can impact their judgments. However, it was not until nearly a century later that empirical data emerged to support Hagan’s admonition. A growing body of literature now shows that forensic examiners are vulnerable to the same confirmation biases that plague us all.

The earliest study to this effect came from Larry Miller (1984), who tested the effects of case information on 12 students who had been trained in questioned document examination (also called handwriting identification) – that is, the ability to assess whether or not two handwriting samples were written by the same author. In a mock forgery investigation, half of these students were told that police had identified a suspect who had been implicated by two eyewitnesses, and then compared this suspect’s handwriting sample against a set of forged documents. The others were told that police had identified three suspects, and then compared each suspect’s sample against the forged documents. In the latter group, all six correctly reported that none of the suspects had committed the forgery; however, in the former group, four of the six incorrectly concluded that the suspect had committed the forgery, and a fifth deemed the evidence inconclusive. To explain this, Miller surmised that examiners who were aware of the eyewitness evidence formed an expectation of guilt that guided their analysis in a self-verifying manner.

These results may be unsurprising, given that the field of handwriting identification has been criticized as among the most subjective and least scientifically-validated of the forensic sciences (e.g., Risinger, Denbeaux, & Saks, 1989). The empirical literature on handwriting examiner performance is scant and rather mixed. While some researchers have argued that
trained handwriting experts possess identification skills superior to those of laypeople (e.g., Inbau, 1939; Kam, Fielding, & Conn, 1997; Kam & Lin, 2003), others noted that these studies exhibited serious methodological flaws (Risinger & Saks, 1996), were carried out by researchers with a vested interest in their outcome (Risinger & Saks, 2003), and nonetheless showed considerable error rates among experts (Risinger, 2007). In any event, the use of handwriting evidence at trial is not uncommon (see Risinger, 2007, for a compendium of cases).

Itiel Dror and colleagues (Dror, Charlton, & Peron, 2006; Dror & Charlton, 2006) have found similar confirmation bias effects among experienced fingerprint examiners. First, Dror et al. (2006) presented five experts with pairs of fingerprints that, unbeknownst to them, they had judged as a match five years earlier. When now told that the prints came from a high profile case of misidentification – thus implying that they would not match – four of the five experts changed their own prior judgments of the same prints, three of whom now concluded that the prints did not match. Dror and Charlton (2006) then had six additional experts examine prints that they had previously judged as either a match or non-match. For some, examiners were also informed of other evidence – namely that the suspect had confessed (implying a match) or that the suspect had a verified alibi (implying a non-match). As a result, 17% of their judgments changed in light of this new information, and four of the six examiners changed at least one of their own prior judgments. Thus, despite the fact that fingerprint experts show highly accurate performance when operating in a vacuum (Tangen, Thompson, & McCarthy, 2011; Ulery, Hicklin, Buscaglia, & Roberts, 2011), their judgments were alarmingly malleable in the face of biasing information (see Dror & Rosenthal, 2008, for a meta-analysis of these two studies).

Confirmation bias has now been demonstrated in other forensic science domains as well. Bieber (2012) described an ongoing research program in which 66 certified arson investigators
analyzed burn patterns to determine whether or not an ignitable liquid had been present. Prior to inspecting the pattern, some investigators were given information which implied that the fire had been started either accidentally or deliberately. When led to believe that the fire was an accident, investigators became more likely to conclude that ignitable fluid was not present relative to those who received no background information or were led to believe that the fire was deliberate.

Nakhaeizadeh, Dror, and Morgan (in press) explored whether confirmation bias affects how forensic anthropologists evaluate skeletal remains. In their study, 41 trained examiners assessed a skeleton for gender, ancestry, and age at death after being given information that implied a particular profile with respect to these characteristics. As expected, examiners gave judgments that were consistent with the implied profile, but factually incorrect. For example, while 69% of control participants correctly identified the skeleton as female, 100% identified it as female when led to believe that it was female, and 72% identified it as male when led to believe that it was male. Similar effects were seen on estimations of age and ancestry.

Thus, biasing information that happens to be accurate can incidentally improve judgment accuracy, while inaccurate information increases the risk of error. Erroneous forensic judgments are costly in the criminal justice system, where they can ultimately lead to wrongful conviction. Kassin, Bogart, and Kerner (2012) analyzed 241 DNA exoneration cases for various evidentiary errors, and found that false confessions were more likely to be accompanied by forensic science errors than by any other type of error. In cases where both were present, forensic science errors were more likely to follow rather than precede the procurement of a false confession.

The authors speculated that these errors were the product of confirmation bias: When aware of a confession, examiners likely infer the suspect’s guilt, which unwittingly renders them more likely to see forensic evidence as incriminating. Confession evidence is uniquely potent
relative to other forms of evidence (Kassin & Neumann, 1997) and thus creates a particularly strong expectation of guilt that affects subsequent decision-making (e.g., Kassin & Wrightsman, 1980; 1981; Kassin & Sukel, 1997). However, confession evidence can be flawed, as false confessions have been uncovered in over a quarter of known DNA exoneration cases (see Kassin et al., 2010). This raises the possibility that forensic examiners who are exposed to a confession which is later shown to be false may produce a judgment that substantiates the confession, but like the confession, is flawed. Indeed, several of the aforementioned empirical studies (Dror & Charlton, 2006; Elaad et al., 1994; Hasel & Kassin, 2009; Kukucka & Kassin, in press) found that knowledge of confession evidence increased the risk of mistaken inculpatory judgments.

A number of real-world wrongful conviction cases also support this hypothesis. In 2004, 13 year-old Tyler Edmonds was convicted of murder and sentenced to life in prison after falsely confessing to abetting his half-sister in the murder of her husband. In his confession, Edmonds described how he and his sister jointly pulled the trigger of a .22 caliber rifle when firing the fatal shot. Despite there being no evidence to corroborate this claim, a state medical examiner -- who had read Edmonds’ confession -- testified that bullet wounds in the victim’s body suggested that the trigger had been pulled by two fingers. After nearly five years in prison, the Mississippi Supreme Court threw out the “scientifically unfounded” testimony of the medical examiner on appeal, and overturned Edmonds’ conviction (Edmonds v. Mississippi, 2007).

Not even DNA – often lauded as the “gold standard” of physical evidence (Lieberman et al., 2008; Lynch, 2003; Saks & Koehler, 2005) -- appears immune to the effects of confirmation bias. Dror and Hampikian (2011) described a gang rape case in which one assailant implicated several other men, and DNA analysts who were aware of the assailant’s testimony concluded that a DNA mixture taken from the victim’s body also implicated the other men. Dror and
Hampikian then sent the DNA evidence from this case, devoid of any contextual information, to 17 independent DNA analysts and found that only one agreed that it implicated the other suspects; four deemed the samples inconclusive, and the remaining 12 concluded that the DNA evidence actually excluded the other men. Thus, the DNA analysts in this case may have been biased to see the DNA mixture as corroborating the assailant’s story.

At least three studies have failed to demonstrate confirmation bias effects on forensic examiners. Kerstholt, Paashuis, and Sjerps (2007) had 12 Dutch officers who were trained in forensic shoe print examination compare photographs of shoe prints from a crime scene against photographs of a suspect’s shoe. For some, these were accompanied by vignettes that contained additional incriminating evidence against the suspect, but this information had no effect on how examiners appraised the shoe prints. Kerstholt et al. (2010) had six Dutch firearms examiners compare pairs of bullets to determine whether they had been fired from the same gun, and were given case information to imply that the two bullets either would or would not match. Again, this manipulation did not have the predicted effect on examiners’ judgments. Finally, Langenburg, Champod, and Wertheim (2009) asked 43 attendees at a conference on fingerprint identification to analyze sets of fingerprints. Some were told of the conclusions reached by another examiner in order to test whether this knowledge would make them likely to concur. However, when given this information, examiners became considerably more likely to judge the fingerprints as inconclusive, and consequently produced zero erroneous judgments.

There are a number of non-mutually exclusive explanations for these failed replications. Kerstholt et al. (2007) speculated that Dutch shoe print examiners may be less subject to bias by virtue of their use of a highly standardized procedure. Broadly speaking, vulnerability to bias may vary between disciplines, techniques, and/or laboratories. Also, each of the Kerstholt studies
featured a small sample size, which restricts the power to detect significant effects (e.g., Cohen, 1990). Finally, Dror (2009) stressed that examiners who are aware of being evaluated as part of a research study are likely to behave in ways that do not accurately reflect how they would perform real-world case work. The notion of demand characteristics (Orne, 1962) posits that research participants may adjust their behavior so as to support or refute what they believe to be the study’s hypothesis. Indeed, many of the examiners in the Langenburg et al. (2009) study later reported that they had ascertained the purpose of the experiment, and thus may have behaved unrealistically. To avoid this problem, it is better to integrate research stimuli into examiners’ routine case work so as to maximize ecological validity (Dror, 2009; Risinger, 2009).

Even without information to suggest guilt or innocence, judgments of forensic evidence may be sensitive to other case information. Dror et al. (2005) found that the emotional intensity of a crime may promote bias in examiners’ judgments. When asked to compare ambiguous pairs of fingerprints, experts were more likely to judge them as a match when they were said to have come from a murder case and were accompanied by gruesome crime scene photos, as opposed to when they came from a petty theft and accompanied by innocuous photos.

Page, Taylor, and Blenkin (2012) speculated that comparisons of bite mark evidence may be susceptible to bias, given that bite mark evidence is often ambiguous and tends to be found in more violent crimes. To test this hypothesis, Osborne, Woods, Kieser, and Zajac (in press) had dental students who had been trained in forensic odontology compare bite marks from a victim’s skin against dental overlays of a suspect’s teeth. When these stimulus pairs were presented with a highly emotive crime scene photo and a subliminal prime (i.e., the words “same” and “guilty”), participants became less likely to judge them as a match, thus producing the opposite effect as
found in Dror et al. (2005). Despite this discrepancy, both studies demonstrated that judgments of forensic evidence can likewise be influenced by non-probative information.

In light of mounting evidence that knowledge of case information can be detrimental to forensic examiners, perhaps the most straightforward remedy is to simply not provide them with such information. Consistent with Rosenthal’s (1978) admonition that scientists be kept “as blind as possible for as long as possible” (p. 1007), many have urged that forensic examiners be denied access to any information that is not essential for their analysis (e.g., Kassin et al., 2013; Loftus & Cole, 2004; Mnookin et al., 2011; Risinger et al., 2002; Stoel, Dror, & Miller, 2014).

To that end, Krane et al. (2008) proposed a method called *sequential unmasking*. Under this method, a qualified “case manager” controls the flow of case information from investigators to examiners. First, the examiner is given a crime-relevant forensic sample, which they analyze in isolation from any information about the suspect or investigation. Once this is complete, the case manager provides the examiner with the suspect’s sample and relevant case information. In so doing, the case manager “masks” (i.e., filters out) any case information that is potentially biasing and not essential for the analysis (see also Dror et al., 2011; Thompson, 2011).

Forensic examiners have raised several objections to the use of masking. As noted earlier, many have stated that unfettered access to case information is essential (Budowle et al., 2009) and beneficial to examiners (Butt, 2013; Elaad, 2013). Others, while acknowledging the biasing impact of case information, have expressed concern over the difficulty of determining what information is and is not “essential” (e.g., Reese, 2012). Finally, some have argued that these protocols would impose a financial burden on forensic laboratories (e.g., Charlton, 2013).

While some forensic examiners “remain stubbornly unwilling to confront and control the problem of bias” (Loftus & Cole, 2004, p. 959), some laboratories have already taken steps to
limit examiners’ exposure to extraneous case information. Found and Ganas (2013) described efforts by the Document Examination Unit of an Australian forensic laboratory to strip cases of any non-essential information before they are given to handwriting examiners. Over three years later, these examiners reported that the changes “have not been complex, overly time-consuming or expensive,” and have produced a number of advantageous outcomes but no negative ones (p. 158). Similar masking procedures have also begun to be implemented among handwriting, DNA, and firearms examiners in the Netherlands (Stoel et al., 2014).

Pressure from investigators. External pressure from investigators may deliberately or inadvertently impel forensic examiners to produce additional inculpatory evidence against a suspect. Saks, Risinger, Rosenthal, and Thompson (2003) explained that investigators often ask examiners who provide exculpatory or inconclusive judgments to re-analyze the same evidence again, which may implicitly communicate that their earlier judgment was inadequate and should be changed. Sometimes this communication takes a less subtle form, as when investigators tell one examiner that another examiner has arrived at a different conclusion, or explicitly tell an examiner what conclusion is expected or desired from them. To remedy these concerns, Risinger et al. (2002) proposed that an Evidence and Quality Control (EQC) officer, who is trained in the relevant forensic discipline, should act as “the sole contact point between the entity requesting the test and the laboratory” (p. 46). The EQC officer could then filter out any suggestive communications from investigators before they can influence the examiner.

Robertson (2010) proposed a similar intermediary model for soliciting expert witnesses. Under this “blind expertise” model, litigants wishing to request an expert opinion must do so through an independent agency that maintains a pool of qualified experts. In turn, the agency impartially selects an expert, provides them with the necessary case materials, and compensates
them regardless of their opinion. Robertson argued that this model would prevent litigants from selecting “hired gun” experts to give favorable testimony, while also minimizing the likelihood of experts becoming biased in favor of the litigant who retains and compensates them (e.g., Murrie et al., 2013). Preliminary data has suggested a third benefit of this model, namely that blinded experts were seen as more credible by jurors than experts who have a clear allegiance to either litigant (Robertson & Yokum, 2012). It is certainly conceivable that a similar EQC model could benefit investigators who solicit forensic examiner opinions.

**Method of evidence presentation.** The manner in which evidence is given to forensic examiners may impact their judgments. For example, fingerprint examiners utilize an Automated Fingerprint Identification System (AFIS) to identify potential matches to unknown fingerprints. To do this, AFIS compares an unknown print against a large digital database of known prints and returns a list of possible matches, presented in descending order of likelihood. Dror, Wertheim, Fraser-Mackenzie, and Walajtys (2012) posited that examiners’ judgments could be biased by this rank order, such that they are more likely to identify those at the top of the list as a match solely by virtue of its serial position. Sure enough, when they varied the order in which AFIS presented the same prints, examiners spent more time analyzing those at the top of the list, and were more likely to misidentify non-matching prints as a match when they were at the top of the list – even when a matching print was also present but lower on the list.

In other disciplines, standard practices for the submission of evidence may inherently compel inculpatory judgments Whitman and Koppl (2010) explained that investigators typically provide a forensic laboratory with two samples to be tested: one crime-relevant sample and one suspect sample. Insofar as examiners infer that investigators have reason to believe that the suspect is guilty (and otherwise would not have submitted their sample for testing), examiners
may hold a base rate expectation that the samples will incriminate the suspect. As Risinger et al. (2002) put it, “investigators do not select suspects or evidence at random, but only those they have some reason to think were connected to the crime” (p. 48).

To counteract this expectation, many researchers (e.g., Cole, 2013; Kassin et al., 2013, Reese, 2012; Whitman & Koppl, 2010) have recommended that examiners instead make their judgments from *evidence lineups*. In an evidence lineup, rather than comparing a crime-relevant sample against one suspect sample, it would instead be compared against an array of samples that includes suspect’s sample as well as several “filler” samples (i.e., known non-matching samples). The task of an examiner using an evidence lineup would be to determine which – if any – of the samples from the array matches the crime-relevant sample.

The standard practice of providing examiners with only one suspect sample is analogous to an eyewitness “showup,” where a witness is shown only one suspect photo and is asked to report whether or not that person is the culprit (Dysart & Lindsay, 2007). Eyewitness showups have likewise been criticized on the grounds that they betray the investigators’ belief that the person in the photo is guilty (Wells et al., 1998), and thus the use of a showup is widely thought to increase the likelihood of an innocent person being misidentified as guilty (Kassin, Tubb, Hosch, & Memon, 2001). Wells et al. (1998) agreed that there is “clear evidence that show-ups are more likely to yield false identifications than are properly-constructed lineups,” and cited several studies to that effect (p. 24). Despite these criticisms, showup identification procedures continue to be widely used by police (e.g., Behrman & Davey, 2001).

A meta-analytic comparison of showups and lineups by Steblay, Dysart, Fulero, and Lindsay (2003) revealed higher overall accuracy for showups, but also found that the two procedures produced different types of systematic errors. Under *target-present* conditions (i.e.,
when the culprit’s photo was present), showups and lineups were equally likely to result in a correct identification of the culprit, but showups produced more incorrect rejections (i.e., judgments that the culprit’s photo was not present, when in fact it was). Under target-absent conditions (i.e., when the culprit’s photo was not present), the use of a lineup rendered witnesses less likely to correctly reject the lineup (i.e., correctly report that the culprit’s photo is not present) and accordingly more likely to make an incorrect identification. These patterns are attributable to the fact that lineups showed higher rates of choosing (i.e., identifying any one of the photos as the culprit) than did showups. That is to say, witnesses who used lineups were more likely to report (either correctly or incorrectly) that the culprit’s photo was present.

Steblay et al. (2003) also looked at the relative likelihood of target-absent showups and lineups to produce “dangerous” misidentifications of an innocent suspect as the culprit. By definition, all incorrect identifications made from target-absent showups were misidentifications of an innocent suspect as the culprit. However, because lineups contain several filler photos (i.e., photos of known-innocent people), some of the incorrect identifications made from target-absent lineups were known errors. When these were excluded, Steblay et al. (2003) found that target-absent showups were more likely than target-absent showups to result in the misidentification of an innocent suspect as guilty – by a margin of 23% to 17%. Thus, they provided some evidence for the belief that showups do in fact increase the risk of false identification.

Risinger et al. (2002) posited that an evidence lineup should function in the same way as an eyewitness lineup, by negating base rate expectations of guilt and reducing “systematic error” relative to showups (p. 49). The use of evidence lineups can also benefit from the vast existing literature on best practices for constructing and administering eyewitness lineups, which can be adapted for use with forensic examiners (Technical Working Group on Eyewitness Evidence,
1999; Wells, 2006; Wells et al., 1998). For example, both the lineup administrator and the examiner should be blind as to which sample in the lineup belongs to the suspect (e.g., Canter, Hammond, & Youngs, 2012); the examiner should be carefully instructed before viewing the lineup, and explicitly reminded that the lineup may or may not contain a matching sample (e.g., Malpass & Devine, 1981; Steblay, 1997); the lineup should include filler samples that are sufficiently similar to the suspect’s sample such that the latter does not stand out (see Wells et al., 1998); and the examiner’s confidence in their judgment should be recorded immediately after the judgment is made and before any feedback is given (e.g., Douglass & Steblay, 2006).

Wells, Wilford, and Smalarz (2013) noted several other advantages of evidence lineups. Because examiners who misidentify a filler sample as a match would have committed a known error, evidence lineups would allow for the estimation of errors rates across a given technique, laboratory, or discipline. By providing insight into their own error rates, examiners’ confidence should become better calibrated to their ability. Similarly, fraudulent or incompetent examiners and flawed methodologies would quickly be exposed by their poor performance, while skilled examiners would be able to definitively establish their expertise. Finally, although some have objected to the use of evidence lineups on the basis of expediency (e.g., Charlton, 2013), Reese (2012) explained that the temporal and financial costs should not be daunting.

Despite widespread support, only one empirical study to date has tested the effect of evidence lineups on judgments of forensic evidence. Miller (1987) asked students who had been trained in forensic hair identification to analyze hair samples in a series of fictitious crimes. For some, they compared a crime scene hair against a (non-matching) hair sample from one suspect (i.e., a target-absent showup). For others, they compared a crime scene hair against an array of five other hairs, none of which matched the crime scene hair (i.e., a target-absent lineup). Miller
found that participants were more likely to misidentify a non-matching hair as a match when using a showup (30%) rather than a lineup (4%). While promising, this study was limited in two important ways. First, because Miller used only target-absent showups and lineups (i.e., only non-matching suspect hairs), it is unclear whether and how the use of evidence lineups affects the ability to correctly identify a matching sample when one is present. Second, some researchers have argued that evidence lineups will mitigate the biasing effect of extraneous case information on examiners’ judgments (e.g., Risinger et al., 2002). However, given that Miller’s students did not receive any information that suggested guilt or innocence, this possibility remains untested.
CHAPTER 4: THE CURRENT STUDIES

In sum, the process whereby confirmation bias contributes to forensic science error is clear. Contextual factors – such as knowledge of extraneous case information or the manner in which evidence is presented – create unconscious expectations that guide evidentiary analyses in a self-verifying manner, rendering even experienced and well-intentioned examiners vulnerable to error. When these mistakes are inculpatory, they are seen as independent corroboration of the suspect’s guilt, thus making the case against the suspect appear stronger than it truly is (Kassin, 2012) and heightening the risk of an innocent person being wrongfully convicted.

In their 2009 report, the NAS lamented the paucity of research “on the important topic of cognitive bias in forensic science... both regarding their effects and methods for minimizing them” (p. 124). The current project was designed to begin filling this gap. In each of three studies, participants assumed the role of mock forensic examiners (specifically, handwriting examiners) in a bank robbery investigation. Participants first read a summary of the case, which for some included information to suggest the suspect’s guilt or innocence. Then, all participants analyzed and provided judgments of handwriting evidence relevant to the investigation.

Study 1 tested whether knowledge of a suspect’s confession – which creates an inference of guilt (Kassin & Neumann, 1997) – would impact judgments of handwriting evidence in a self-fulfilling manner, such that participants who were previously informed of the confession would judge the handwriting evidence as more incriminating than those who were not. Study 1 also tested three potential moderators of this bias – namely, the “elasticity” (i.e., similarity) of the handwriting evidence (Ask et al., 2008), asking participants to re-analyze the same evidence (Saks et al., 2003), and participants’ dispositional Need for Cognition (Cacioppo & Petty, 1982).
Study 2 tested whether the use of an evidence lineup would lessen the impact of biasing case information on mock handwriting examiners’ judgments relative to the standard evidence “showup”. Study 2 also compared the effects of showups and lineups on judgment accuracy. In so doing, Study 2 explored whether evidence lineups affect accuracy relative to showups in the same way as eyewitness lineups, as some have surmised (e.g., Risinger et al., 2002).

Study 3 was later added to test whether varying the manner in which evidence lineups were presented (i.e., simultaneously versus sequentially; Lindsay & Wells, 1985) would have a similar effect on mock handwriting examiners’ judgments as it does on eyewitness judgments in terms of accuracy and systematic error (e.g., Steblay, Dysart, & Wells, 2011).
CHAPTER 5: STUDY ONE

Method

In Study 1, participants serving as mock forensic handwriting examiners were asked to evaluate handwriting samples in the context of a bank robbery investigation. Participants read a case summary in which the suspect either maintained his innocence or confessed but later recanted his confession. Then, they compared two handwriting samples – the suspect’s Miranda waiver form and the robbery note that the perpetrator gave to the bank teller. These two samples were either high or low in pre-rated similarity (and, unbeknownst to participants, were written by different authors). After providing judgments of these samples, participants were asked to revisit the case summary and were given the opportunity to affirm or revise their previous judgments. Study 1 thus employed a 2 (Confession: Present vs. Absent) X 2 (Handwriting Similarity: Low vs. High) X 2 (Time: Time 1 vs. Time 2) mixed factorial design.

Participants and Design

A sample of 115 individuals was obtained using Amazon Mechanical Turk (mTurk), an online marketplace service that allows for the recruitment and compensation of individuals to complete online tasks (see Mason & Suri, 2012). The use of mTurk in behavioral research has proliferated in recent years, as researchers have favorably evaluated the service for providing access to diverse participant pools and for permitting the efficient and inexpensive collection of high quality data (e.g., Buhrmester, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013; Paolacci, Chandler, & Ipeirotis, 2010).

Although mTurk allows for the recruitment of participants worldwide, participation in the current study was restricted to U.S. residents only. Each participant was randomly assigned to one of four cells produced by the two between-subjects factors of the 2 (Confession: Present vs. Absent) X 2 (Handwriting Similarity: Low vs. High) X 2 (Time: Time 1 vs. Time 2) mixed
design. Five participants (4.35%) were later excluded after having responded incorrectly to a manipulation check question which asked whether or not the suspect had confessed to the crime, leaving a final sample of $N = 110$ for all analyses (unless otherwise noted).

Participants were predominantly female (66.06%) and had a mean age of 42.02 ($SD = 15.32$; Range = 19 – 73). All participants reported that they are current U.S. citizens, and the sample included at least one resident from 30 of the 50 U.S. states. With respect to race, most self-identified as White (79.82%), with others identifying as Black (8.26%), Hispanic (5.50%), Asian (3.67%), and multi-racial (2.75%). In terms of education level, 35.78% of participants did not hold a college degree, 22.02% held a two-year (i.e., Associates) degree, 29.36% held a four-year (i.e., Bachelors) degree, and 12.84% held a graduate-level (i.e., Masters or Doctoral) degree.

**Procedure**

Participants completed the study using an online survey website. In exchange for their participation, they received a $0.50 credit to their Amazon mTurk account; this rate of compensation is commensurate with mTurk tasks of comparable length (Mason & Suri, 2012).

After giving informed consent (see Appendix A), participants completed an 18-item abbreviated version of the Need for Cognition (NFC) scale (Cacioppo, Petty, & Kao, 1984; see Appendix B). Participants also answered a series of basic demographic questions in which they reported their age, gender, race, U.S. citizenship, state of residence, and level of educational attainment. In order to minimize any priming effects from having completed the NFC scale, several filler questions were included but not analyzed; these included questions about political affiliation, marital status, handedness, and international travel (see Appendix C).

Participants were then told that for the remainder of the study they would assume the role of a handwriting identification expert. Instructions explained that the role of a handwriting expert
is to compare handwriting samples and offer opinions as to whether the samples were written by the same person, and that because their opinions can be useful in solving crimes, they are often enlisted by police investigators to assist with investigations (see Appendix D).

Next, participants read a simulated memo from a police Sergeant that requested their assistance with an ongoing armed robbery investigation. The memo explained that they would be given information about an actual investigation and would be asked to review the case, examine relevant handwriting evidence, and report their opinions of the evidence back to police.

This memo was followed by one of two summaries of a bank robbery investigation, modeled after the case of *U.S. v. Hines* (1999; see Appendix E). In each, an armed man gave a handwritten note to a bank teller and escaped with a large sum of cash. Police then apprehended a suspect (Johanna Hines) who matched the teller’s general description of the perpetrator, and brought him to the police station, where he handwrote and signed a waiver of his *Miranda* rights and was questioned by police for three hours. By random assignment, half of the participants were told that Hines gave a detailed confession to the robbery, which he later recanted, claiming that he was coerced by police (*confession-present* condition). The other half were told that Hines maintained his innocence throughout the police interview (*confession-absent* condition).

After reading either case summary, participants were shown two items of handwriting evidence, presented side-by-side on the screen. These included: (1) the note that the perpetrator handed to the bank teller, and (2) the *Miranda* waiver written by Hines prior to being questioned by police. By random assignment, participants received handwriting stimuli that were either high (*high similarity* condition) or low (*low similarity* condition) in objective similarity, based on earlier pilot testing by Kukucka and Kassin (in press; see Appendix F). Participants were given an unlimited amount of time to compare the samples. During this time, they rated the similarity
of the two samples, gave their opinion as to whether or not they had been authored by the same individual (i.e., were a “match”), and indicated their confidence in this latter judgment.

After submitting these judgments, participants received a second memo from the police department, which asked them to review the facts of the case again, stating that experts make better judgments when given more time to review the evidence (see Appendix G). Participants were then re-presented with the same case summary and the same pair of handwriting samples, and were asked to provide a final set of the same judgments that they had given previously. This was designed to simulate real-world scenarios in which investigators ask forensic examiners to re-evaluate the same evidence, which may implicitly communicate that their first judgment was undesirable and should be changed (Saks et al., 2003).

Next, participants answered two self-report items regarding the extent to which they felt that their judgments of the handwriting evidence were influenced by the handwriting stimuli and by the facts of the case, respectively. Finally, they completed a comprehension test to ensure that they read, understood, and accurately recalled the details of the case summary (see Appendix H), and all were fully debriefed with respect to the purpose of the study.

Materials

Need for cognition. Participants completed an abbreviated version of the Need for Cognition scale (NFC; Cacioppo et al., 1984; see Appendix B), which features 18 of the 34 items from the original NFC scale developed by Cacioppo and Petty (1982). Each item consists of a statement (e.g., “I would prefer complex problems to simple problems”) for which participants indicate the extent to which they agree or disagree with the statement, using a scale that ranges from −4 (very strongly disagree) to +4 (very strongly agree). Nine of these items were reverse-
scored and responses to all 18 items were then summed, which resulted in total scores that could range from -72 to +72, with higher scores corresponding to higher levels of NFC.

Four participants (3.64%) failed to answer one or more items on the NFC scale, and thus total NFC scores could not be computed for these individuals. The remaining 106 participants produced a mean total NFC score of 26.76 ($SD = 27.41$; Range $= -64 – 72$) and their responses showed strong internal consistency (Cronbach’s $\alpha = .95$). Participants were later categorized as either high ($n = 53$) or low ($n = 53$) in NFC on the basis of a median split ($Med = 30.50$).

**Case summary.** We utilized a case summary that was developed and used by Kukucka and Kassin (in press) and based on *U.S. v. Hines* (1999; see Appendix E). In *Hines*, District Judge Nancy Gertner ruled that a questioned document examiner who was proffered as an expert witness could not express any definitive conclusion as to whether two handwriting samples were a “match,” as such a conclusion would have no scientific basis. Instead, Judge Gertner permitted the expert to point out similarities and differences between the two samples, and allowed jurors to draw their own inferences with respect to authorship. Many details of the case summary were modeled after the original case, including the race and gender of the perpetrator and bank teller, the name and location of the bank, the amount of money stolen, and the bank teller’s description of the perpetrator and subsequent eyewitness identification of Hines.

In the summary, a young African-American man gave a handwritten note to a bank teller which read, “*I have a gun. Keep quiet or I will shoot you. Give me all your cash!*” The robber opened his coat to reveal a concealed handgun and fled with over $10,000 in cash. Police arrived and interviewed the bank teller, who described the robber as a tall Black male wearing a heavy coat and jeans. The summary also included a time-stamped, low-quality surveillance photo which was consistent with the teller’s generic description of the perpetrator.
Approximately a half-hour after the robbery, police stopped a speeding vehicle in the vicinity of the bank and found that the driver, Johanna Hines, matched the description given by the bank teller. The officer reported that Hines appeared nervous while being interviewed, but a search of his vehicle revealed neither a gun nor the stolen cash. When shown a photo lineup of six men who fit the description she had given, the bank teller identified Hines as the culprit but admitted that she was not confident in her identification. Hines was then picked up by police and brought to the station for questioning. Prior to being interviewed, Hines produced a handwritten waiver of his Miranda rights which read, “I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.” Hines was then questioned for three hours.

**Confession manipulation.** Participants in the *confession-present* condition were told that after three hours, Hines confessed to the robbery. They were shown a type-written confession statement, in which Hines explained that he robbed the bank because he was out of work and in debt. The statement also included a description of the bank teller, and described how he bought the gun, wrote a note for the teller, disguised his appearance, and hid the gun and stolen cash in a dumpster after fleeing the bank. This confession statement was shown on police department letterhead and signed illegibly. After meeting with a lawyer, Hines recanted his confession claiming that he was coerced by police, and pleaded not guilty to armed robbery.

Participants in the *confession-absent* condition were told that Hines maintained his innocence throughout the three-hour interview. When asked for his whereabouts during the robbery, Hines told police that he was eating breakfast alone at a local restaurant. To ensure that the level of detail was equivalent between the two conditions, participants were shown a type-written and signed denial statement on letterhead, in which Hines gave the name and address of
the restaurant and described the cashier, what he had eaten, and what he did while he was there. Despite his denials, Hines was charged with armed robbery, to which he pleaded not guilty.

**Handwriting similarity manipulation.** Participants were shown one of two pairs of handwriting samples, each of which consisted of a robbery note and a *Miranda* waiver that had been written by different authors. By random assignment, participants received either the high similarity pair or the low similarity pair (see Appendix F). Pilot testing by Kukucka and Kassin (in press) found that the high similarity pair was rated as more similar and was more often judged as a match than the low similarity pair (*ps* < .0001).

**Dependent measures.** Participants made three judgments concerning the handwriting samples. First, they rated the similarity of the handwriting in the robbery note (written by the perpetrator) and in the *Miranda* waiver (written by Hines, the suspect), using a scale that ranged from 1 (not at all similar) to 10 (very similar). Second, they gave a trichotomous judgment as to whether they believed that two notes were authored by the same individual (i.e., a “match”), with response options of “yes,” “no,” and “cannot be determined” (CBD). Third, participants who gave a definitive match judgment of “yes” or “no” indicated their confidence in this judgment on a scale from 1 (not at all confident) to 10 (very confident). Those who responded that a match “cannot be determined” were assigned a confidence rating of zero.

After revisiting the case summary, participants provided these same three handwriting judgments for a second time. Two additional items followed the second set of handwriting judgments. On a scale of 1 (not at all) to 10 (very much), participants self-reported the extent to which they believed that their judgments of the handwriting evidence were influenced (a) by the handwriting samples themselves, and (b) by the facts of the case.
Comprehension test. Participants answered five multiple-choice questions to ensure that they read, understood, and recalled the content of the case summary (see Appendix H). This included one item which asked whether the suspect had previously confessed to the robbery; five participants (4.35%) were later excluded after answering this item incorrectly.

Hypotheses

H1: Based on research showing that confessions can color people’s perceptions and judgments (e.g., Kassin et al., 2013), I predicted that participants who were told that the suspect had confessed to the robbery would rate handwriting samples from the suspect and robber as more similar, and would misjudge them as a “match” (i.e., as having been authored by the same person) more often and more confidently, than those in the no-confession control condition.

H2: In light of research on the ‘elasticity’ of criminal evidence (e.g., Ask et al., 2008), I predicted that the pre-rated similarity of the handwriting samples would moderate the biasing effect of the confession, such that the confession would have a greater impact on judgments of handwriting samples that were high rather than low in similarity.

H3: In line with claims that re-examination of evidence can produce bias (e.g., Saks et al., 2003), I predicted that the biasing effect of the confession would be exacerbated by asking participants to revisit the case summary, such that the confession would have a greater impact on judgments of handwriting samples at Time 2 than at Time 1.

H4: Given research to suggest that Need for Cognition (NFC; Cacioppo & Petty, 1982) moderates susceptibility to confirmation bias (e.g., Kassin et al., 1990), I predicted that the biasing effect of the confession on handwriting judgments would be greater among participants who exhibited a high rather than low dispositional level of NFC.
H5: Consistent with the notion that people maintain an “illusion of objectivity” when their cognition is biased (Kunda, 1990), I predicted that participants who were aware of the confession would under-value its impact, such that they would report that their handwriting judgments were influenced more by the handwriting itself than by the facts of the case.
CHAPTER 6: STUDY ONE RESULTS

Effects of Confession, Similarity, and Time

Hypothesis 1 posited that participants who were told of the suspect’s confession would rate handwriting samples from the perpetrator and suspect as more similar relative to those who were not told of the confession. Hypothesis 2 posited that this effect would be stronger for handwriting samples that were high versus low in similarity, and Hypothesis 3 posited that this effect would be more pronounced at Time 2 than at Time 1. Across all conditions and time, participants produced a mean similarity rating of 4.43 (SD = 2.67) on a 10-point scale.

To address these hypotheses, a 2 (Confession: Present vs. Absent) X 2 (Handwriting Similarity: Low vs. High) X 2 (Time: Time 1 vs. Time 2) mixed ANOVA was performed on similarity ratings. The predicted main effect of Confession was found, $F(1,106) = 4.26, p = .041, d = 0.36$ [95% CI: 0.01 – 0.71], such that participants who were told of the suspect’s confession ($M = 4.95, SD = 2.65$) judged the handwriting samples as more similar than those who were not ($M = 4.00, SD = 2.63$). A main effect of Similarity confirmed the pilot testing of Kukucka and Kassin (in press), $F(1,106) = 3.97, p = .049, d = 0.37$ [95% CI: 0.02 – 0.72], with the high similarity pair ($M = 4.92, SD = 2.83$) rated as more similar overall than the low similarity pair ($M = 3.95, SD = 2.43$). No main effect of Time was found, $F(1,106) = 0.14, p = .712, d = 0.03$ [95% CI: -0.18 – 0.24], as the mean similarity rating did not change between Time 1 ($M = 4.41, SD = 2.65$) and Time 2 ($M = 4.45, SD = 2.72$). None of the two-way interactions was significant – including Confession X Time, $F(1,106) = 0.09, p = .769, \eta^2_p = .00$, Confession X Similarity, $F(1,106) = 1.88, p = .173, \eta^2_p = .02$, and Similarity X Time, $F(1,106) = 1.76, p = .188, \eta^2_p = .02$ – nor was the three-way interaction significant, $F(1,106) = 0.89, p = .347, \eta^2_p = .01$. 
Despite the absence of a Confession X Similarity interaction in the three-way mixed ANOVA, I performed follow-up 2 (Confession) X 2 (Time) ANOVAs within each level of Similarity to test my a priori prediction that Similarity would moderate the effect of Confession on similarity ratings (i.e., Hypothesis 2; see Figure 1). Among High Similarity pairs, Confession had no effect, $F(1,53) = 0.19, p = .664, d = 0.12$ [95% CI: -0.41 – 0.64], such that similarity ratings were equal when the confession was present ($M = 5.10, SD = 3.01$) and absent ($M = 4.77, SD = 2.69$). Neither the effect of Time, $F(1,53) = 0.60, p = .441, d = 0.12$ [95% CI: -0.18 – 0.42], nor the Confession X Time interaction, $F(1,53) = 1.02, p = .318, \eta^2_p = .02$, was significant.

Among Low Similarity pairs, however, a main effect of Confession emerged, $F(1,53) = 7.93, p = .007, d = 0.72$ [95% CI: 0.29 – 1.14], such that similarity ratings were higher when the confession was present ($M = 4.81, SD = 2.28$) than when it was not ($M = 3.17, SD = 2.31$). Neither the effect of Time, $F(1,53) = 1.16, p = .287, d = 0.15$ [95% CI: -0.15 – 0.45], nor the Confession X Time interaction, $F(1,53) = 0.17, p = .682, \eta^2_p = .00$, was significant.

Hypotheses 1-3 likewise predicted that the confession would impact match judgments and their associated confidence ratings, and that this effect would be moderated by pre-tested handwriting similarity and time. To allow for a more sensitive test of match judgments, I combined these two dependent measures into a single match-confidence composite score for each participant at both Time 1 and Time 2. Judgments that the samples were a match were coded as +1, judgments that they were not a match were coded as -1, and judgments of “cannot be determined” were coded as 0. I then computed the product of these values and participants’ confidence ratings, thereby producing match-confidence composite scores that could range from -10 (highly confident “non-match” judgment) to +10 (highly confident “match” judgment).

Across all conditions and time, the mean match-confidence composite score was -3.80 ($SD = \ldots$)
6.55); the negative value of this mean reflects the fact that, across all conditions, samples were judged as a non-match 68.18% of the time and as a match only 19.55% of the time.

A 2 (Confession) X 2 (Handwriting Similarity) X 2 (Time) mixed ANOVA was performed on these match-confidence composite scores. A marginal effect of Confession was found, $F(1,106) = 3.62, p = .060, d = 0.33 \ [95\% \ CI: -0.52 - 1.19]$, such that participants who were told of the confession ($M = -2.63, SD = 6.97$) had higher composite scores than those who were not ($M = -4.78, SD = 6.03$). A main effect of Similarity was also found, $F(1,106) = 4.21, p = .043, d = 0.38 \ [95\% \ CI: -0.47 - 1.22]$, with high similarity samples ($M = -2.60, SD = 7.13$) producing higher composite scores than low similarity samples ($M = -5.01, SD = 5.69$). No effect of Time was found, $F(1,106) = 0.08, p = .773, d = 0.04 \ [95\% \ CI: -0.17 - 0.25]$, as composite scores did not change from Time 1 ($M = -3.73, SD = 6.57$) to Time 2 ($M = -3.88, SD = 6.55$). None of the two-way interactions was significant – including Confession X Time, $F(1,106) = 1.05, p = .307, \eta^2_p = .01$, Confession X Similarity, $F(1,106) = 0.51, p = .476, \eta^2_p = .01$, and Similarity X Time, $F(1,106) = 3.11, p = .081, \eta^2_p = .03$ – nor was the three-way interaction significant, $F(1,106) = 2.63, p = .108, \eta^2_p = .02$. Thus, the observed pattern of results for match-confidence composite scores was virtually identical to that of similarity ratings; indeed, these measures proved to be strongly correlated, $r(218) = 0.79, p < .001$.

**Time 1 Judgments**

Separate analyses were conducted on judgments at Time 1 and Time 2 to better understand the observed effects of Confession and Similarity. At Time 1, participants produced an overall mean similarity rating of 4.41 ($SD = 2.65$); 20.00% judged the samples as a match, 68.18% judged them as a non-match, and 11.82% selected “cannot be determined” (CBD), producing an overall mean match-confidence composite score of -3.73 ($SD = 6.57$).
A 2 (Confession) X 2 (Similarity) ANOVA on similarity ratings at Time 1 revealed a marginal effect of Confession, \( F(1,106) = 3.73, p = .056, d = 0.35 \) [95% CI: -0.14 – 0.83], such that participants who were told of the confession judged the samples as more similar \( (M = 4.90, \ SD = 2.63) \) than those who were not \( (M = 4.00, \ SD = 2.61) \). An effect of Similarity also emerged, \( F(1,106) = 5.53, p = .021, d = 0.46 \) [95% CI: -0.02 – 0.94], with high similarity pairs judged as more similar \( (M = 5.00, \ SD = 2.87) \) than low similarity pairs \( (M = 3.82, \ SD = 2.27) \). The Confession X Similarity interaction was not significant, \( F(1,106) = 2.65, p = .107, \eta^2_p = .02 \).

A series of chi-square tests was used to test the effects of Confession and Similarity on trichotomous match judgments. The presence of a confession had no overall effect on match judgments, \( \chi^2(2) = 0.97, p = .615 \), Cramér’s \( V = .09 \). When a confession was present, 24.00% judged the samples as a match, 64.00% judged them as a non-match, and 12.00% gave CBD judgments. When a confession was absent, 16.67% judged them as a match, 71.67% judged them as a non-match, and 11.67% gave CBD judgments. Follow-up tests for moderation revealed that Confession did not impact match judgments of either the high similarity, \( \chi^2(2) = 0.17, p = .919, \) Cramér’s \( V = .05 \), or the low similarity, \( \chi^2(2) = 3.53, p = .171 \), Cramér’s \( V = .25 \), pair.

I then excluded participants who gave CBD judgments to test whether knowledge of the confession affected the likelihood of definitively concluding that the samples were a match or non-match. Among those who gave match or non-match judgments, the effect of Confession remained non-significant, \( \chi^2(1) = 0.97, p = .325, \phi = .10 \), such that participants were equally likely to judge the samples as a match when a confession was present \( (27.27\%) \) or absent \( (18.87\%) \). Follow-up tests for moderation by Similarity found that Confession had no effect on definitive match judgments of the high similarity pair, \( \chi^2(1) = 0.01, p = .905, \phi = .02 \). However, there was a marginal effect of Confession for the low similarity pair, \( \chi^2(1) = 3.41, p = .065, \phi = \).
such that participants were more likely to judge the low similarity pair as a match when the Confession was present (20.83%) than when it was absent (3.85%).

Lastly, a 2 (Confession) X 2 (Similarity) ANOVA was performed on match-confidence composite scores at Time 1, which confirmed a main effect of Similarity, $F(1,106) = 6.17, p = .015, d = 0.49$ [95% CI: -0.70 – 1.68], such that composite scores were higher for high similarity pairs ($M = -2.18$, $SD = 7.27$) than low similarity pairs ($M = -5.27$, $SD = 5.44$). Neither the effect of Confession, $F(1,106) = 2.37, p = .126, d = 0.27$ [95% CI: -0.96 – 1.50], nor the Confession X Similarity interaction, $F(1,106) = 1.33, p = .251, \eta^2_p = .01$, reached significance.

**Time 2 Judgments**

Across all conditions at Time 2, participants produced an overall mean similarity rating of 4.45 ($SD = 2.72$); 19.09% judged the samples as a match, 68.18% judged them as a non-match, and 12.73% gave CBD judgments. Accordingly, participants produced a mean match-confidence composite score of -3.88 ($SD = 6.55$) at Time 2.

A 2 (Confession) X 2 (Similarity) ANOVA on similarity ratings at Time 2 again showed a main effect of Confession, $F(1,106) = 4.05, p = .047, d = 0.38$ [95% CI: -0.12 – 0.87], such that similarity ratings were higher when the confession was present ($M = 5.00$, $SD = 2.69$) than absent ($M = 4.00$, $SD = 2.67$). Unexpectedly, the main effect of Similarity was not significant at Time 2, $F(1,106) = 2.19, p = .142, d = 0.29$ [95% CI: -0.21 – 0.79], with the high ($M = 4.84$, $SD = 2.80$) and low similarity ($M = 4.07$, $SD = 2.60$) pairs being rated as equally similar. The Confession X Similarity interaction, $F(1,106) = 1.02, p = .316, \eta^2_p = .01$, was not significant.

A series of chi-square analyses explored the effects of Confession and Similarity on trichotomous match judgments. As at Time 1, the presence of the confession had no overall effect on these judgments, $\chi^2(2) = 4.46, p = .108$, Cramér’s $V = .20$. When the confession was
present, 26.00% judged the samples as a match, 58.00% judged them as a non-match, and 16.00% gave CBD judgments. When the confession was absent, 13.33% judged them as a match, 76.67% judged them as a non-match, and 10.00% gave CBD judgments. Follow-up tests for moderation revealed that Confession did not impact match judgments of either the high, $\chi^2(2) = 2.67$, $p = .263$, Cramér’s $V = .22$, or low similarity, $\chi^2(2) = 3.04$, $p = .219$, Cramér’s $V = .24$, pair.

As I did at Time 1, I then excluded CBD judgments and performed a second set of chi-square tests using only those participants who gave definitive match or non-match judgments. The presence of a confession had a marginal effect on definitive judgments, $\chi^2(1) = 3.60$, $p = .058$, $\phi = .19$, such that the samples were more often judged as a match when the confession was present (30.95%) than when it was not (14.81%). Follow-up tests for moderation by Similarity revealed no significant effect of Confession on definitive match judgments of either the high, $\chi^2(1) = 1.34$, $p = .246$, $\phi = .17$, or low similarity, $\chi^2(1) = 3.02$, $p = .082$, $\phi = .25$, pair.

Lastly, a 2 (Confession) X 2 (Similarity) ANOVA was performed on match-confidence composite scores at Time 2. A main effect of Confession emerged, $F(1,106) = 1.90$, $p = .053$, $d = 0.40$ [95% CI: -0.80 – 1.59], such that composite scores were higher when a confession was present ($M = -2.50$, $SD = 7.07$) rather than absent ($M = -5.03$, $SD = 5.91$). Neither the effect of Similarity, $F(1,106) = 2.11$, $p = .149$, $d = 0.27$ [95% CI: -0.94 – 1.48], nor the Confession X Similarity interaction, $F(1,106) = 0.05$, $p = .825$, $\eta^2_p = .00$, reached significance.

**Change Scores**

The observed patterns of results at Time 1 and Time 2 were largely identical. To explore changes in handwriting judgments over time, I computed three change scores for each participant which would allow for more precise tests of Hypothesis 3. First, I calculated changes in similarity ratings of the same handwriting pair from Time 1 to Time 2. On average, similarity
ratings increased by 0.05 (SD = 1.53) from Time 1 to Time 2; 70.00% of similarity ratings did not change. Second, using the aforementioned coding scheme for match judgments (i.e., +1 for match, 0 for CBD, -1 for non-match), I noted whether match judgments became more positive, more negative, or did not change over time. Overall, 88.18% did not change their match judgment, while 5.45% and 6.36%, respectively, gave judgments that were more positive or more negative at Time 2 than at Time 1. Third, I calculated changes in match-confidence composite scores from Time 1 to Time 2. On average, composite scores decreased by 0.15 (SD = 3.73) from Time 1 to Time 2; 63.63% of composite scores did not change.

A 2 (Confession) X 2 (Similarity) ANOVA on changes in similarity rating showed no main effect of Confession, F(1,106) = 0.09, p = .769, d = 0.07 [95% CI: -0.22 – 0.35], such that changes in similarity rating from Time 1 to Time 2 were equivalent whether the confession was present (M = 0.10, SD = 0.91) or absent (M = 0.00, SD = 1.91). Neither the main effect of Similarity, F(1,106) = 1.76, p = .188, d = 0.28 [95% CI: -0.01 – 0.56], nor the Confession X Similarity interaction, F(1,106) = 0.89, p = .347, η²_p = .01, reached significance.

To test the effect of Confession on the direction and frequency of changes in match judgments, the scarcity of such changes precluded the use of a Pearson chi-square test, as four of the six cells in the 2 X 3 table had expected values below n = 5. Instead, descriptive results are provided, along with a probability estimate from the Freeman-Halton test (i.e., an adaptation of Fisher’s exact test for 2 X 3 tables). When a confession was present, 8.00% of participants showed a positive change in match judgment over time (i.e., movement toward a match judgment), while only 2.00% showed a negative change. When a confession was not present, 3.33% showed positive change over time, while 10.00% showed a negative change. Though in the predicted direction, this effect did not achieve statistical significance, p = .139.
Finally, a 2 (Confession) X 2 (Similarity) ANOVA on changes in match-confidence composite scores revealed no effect of Confession, \( F(1,106) = 1.05, p = .307, d = 0.21 \ [95\% \text{ CI: } -0.48 \text{ – } 0.90] \), nor an effect of Similarity, \( F(1,106) = 3.11, p = .081, d = 0.37 \ [95\% \text{ CI: } -0.31 \text{ – } 1.06] \), nor a Confession X Similarity interaction, \( F(1,106) = 2.63, p = .108, \eta^2_p = .02 \).

**Need for Cognition**

Hypothesis 4 posited that Need for Cognition (NFC) would moderate the effect of confession, such that it would have a greater impact on the judgments of high-NFC participants. To test this hypothesis, NFC score (dichotomized at the median) was included as a between-subjects factor in a 2 (Confession) X 2 (Similarity) X 2 (NFC Score: Low vs. High) X 2 (Time) mixed ANOVA. NFC Score had no main effect on similarity ratings, \( F(1,98) = 0.34, p = .563, d = 0.18 \ [95\% \text{ CI: } -0.18 \text{ – } 0.53] \), nor did it interact with Confession, \( F(1,98) = 1.56, p = .214, \eta^2_p = .02 \), Similarity, \( F(1,98) = 0.77, p = .382, \eta^2_p = .01 \), or Time, \( F(1,98) = 0.06, p = .804, \eta^2_p = .00 \).

It should be noted that this test was severely underpowered (see Cohen, 1990) due to the limited size of the sample and the inclusion of three between-subjects factors, which resulted in cell sizes as low as \( n = 8 \). For example, the observed level of power associated with the predicted Confession X NFC Score interaction was \( 1 - \beta = .24 \), indicating that the test would fail to detect a significant interaction 76\% of the time. When Similarity was removed as a between-subjects factor, the Confession X NFC Score interaction remained non-significant, \( F(1,102) = 2.59, p = .111, \eta^2_p = .03 \), but the observed degree of power was again quite low (\( 1 - \beta = .36 \)).

**Self-Reported Influence**

Using a 10-point scale, participants self-reported the extent to which their handwriting judgments were influenced by the handwriting itself and by the facts of the case. Hypothesis 5
posited that they would believe that their judgments were influenced more by the handwriting than by the case facts, suggesting an “illusion of objectivity” (Kunda, 1990).

A 2 (Confession) X 2 (Similarity) X 2 (Source of Influence: Handwriting vs. Case Facts) mixed ANOVA with repeated measures on the third factor revealed the predicted main effect of Source, $F(1,106) = 31.58, p < .001, d = 0.87$ [95% CI: 0.54 – 1.21], such that participants believed that their handwriting judgments were influenced more by the handwriting samples themselves ($M = 8.17, SD = 2.26$) than by the facts of the case ($M = 5.95, SD = 2.82$).

This was qualified by a Confession X Source interaction, $F(1,106) = 5.84, p = .017, \eta^2_p = .05$ (see Figure 2). Simple effects tests showed that when the confession was absent, participants felt more influenced by the handwriting ($M = 8.52, SD = 2.08$) than by the case facts ($M = 5.48, SD = 2.91$), $t(59) = 6.11, p < .001, d = 0.79$ [95% CI: 0.50 – 1.08]. When the confession was present, participants again felt more influenced by the handwriting ($M = 7.76, SD = 2.42$) than by the case facts ($M = 6.52, SD = 2.63$), but the magnitude of this difference was weaker, $t(49) = 2.17, p = .035, d = 0.31$ [95% CI: -0.01 – 0.62]. Stated differently, participants felt equally influenced by the handwriting regardless of whether the confession was present, $t(108) = 1.76, p = .081, d = 0.34$ [95% CI: -0.07 – 0.76], but felt marginally more influenced by the case facts when the confession was present than when it was not, $t(108) = -1.94, p = .055, d = -0.38$ [95% CI: -0.89 – 0.14]. Neither the Similarity X Source interaction, $F(1,106) = 0.71, p = .403, \eta^2_p = .01$, nor the three-way interaction, $F(1,106) = 1.87, p = .174, \eta^2_p = .02$, was significant.
CHAPTER 7: STUDY ONE DISCUSSION

As predicted, mock handwriting examiners in Study 1 who were told of a suspect’s confession judged two non-matching handwriting samples as more similar, relative to those who were told that the suspect denied guilt. Our findings thus replicated the findings of Kukucka and Kassin (in press) with respect to the impact of confession evidence on handwriting judgments, and further underscore the concerns of many researchers -- as well as the National Academy of Sciences (2009) -- over the potential for contextual bias to influence forensic science analyses. It is worth reiterating that neither of the pairs of handwriting samples used in Study 1 was, in fact, written by the same author, and therefore every erroneous “match” judgment was of the type that would implicate an innocent person. Given the frequency with which forensic science errors are known to follow the procurement of a false confession (Kassin et al., 2012), the findings of Study 1 further raise concerns over allowing forensic science examiners access to confession evidence that is not germane to their circumscribed forensic analysis.

It is important to note that knowledge of the suspect’s confession impacted handwriting judgments even though the confession manipulation was relatively weak, insofar as the suspect recanted it and claimed to have been coerced by police. This effect is consistent with prior research showing that confession evidence impacts decision-making even when the confession is seen as involuntary (e.g., Kassin & Sukel, 1997; Kassin & Wrightsman, 1980). In all likelihood, the confession would have had a far stronger impact had participants not been told that it was recanted before comparing the handwriting samples. Indeed, given that participants learned of the recantation immediately prior to viewing the handwriting samples, the recantation was likely highly salient, which may have attenuated the confession effect. Future studies should manipulate the presence and/or timing of the recantation to explore these possibilities.
Participants showed some insight into the effect of case facts on their handwriting judgments, as those who were told of the confession felt somewhat more influenced by the case facts than who were not. However, although those who were told of the confession were clearly impacted by it, they maintained an “illusion of objectivity” with respect to their handwriting judgments, believing them to be more a product of the handwriting evidence itself than of the case facts (Kunda, 1990). It is important to note that we did not ask participants the extent to which they were influenced by the confession. Prior research has shown that confession evidence impacts decision-making even when decision-makers claim to have been unaffected by it (e.g., Kassin & Sukel, 1997). We also did not caution our participants against allowing the case facts to affect their handwriting judgments. Indeed, even those who were not told of the confession claimed to be moderately influenced by the facts of the case. Adding a cautionary instruction might inhibit participants’ willingness to report being influenced by case facts (e.g., Orne, 1962), but it would be unlikely to ameliorate bias, given research showing that mere awareness of the possibility of bias is not sufficient to reduce bias (e.g., Wilson et al., 1996).

I found some support for the prediction that the pre-rated similarity of the handwriting samples would moderate the impact of the confession -- but, this moderation was in the opposite direction as predicted, such that the confession increased similarity ratings of the low-similarity pair but not of the high-similarity pair. This unexpected finding may seem difficult to reconcile with basic research indicating that confirmation bias is more likely to distort the perception of stimuli that are ambiguous rather than unambiguous (e.g., Fisher, 1968; Kunda, 1990). Indeed, previous studies have shown that confession evidence impacted judgments of inconclusive, but not conclusive, polygraph charts (Elaad et al., 1994) and judgments of “difficult” fingerprint pairs more so than “not difficult” pairs (Dror & Charlton, 2006). With this in mind, I presumed
that the high-similarity pair would prove more ambiguous and therefore judgments of it would be more malleable, whereas the low-similarity pair would provide too little bottom-up input to justify a “match” judgment even in the face of a confession (Darley & Gross, 1983).

In order to identify low- and high-similarity pairs of handwriting samples for Study 1, I returned to a pilot study by Kukucka and Kassin (in press) in which participants made context-free judgments of handwriting sample pairs. From these, I selected two pairs that significantly differed in terms of their mean similarity ratings and match judgment probabilities, under the assumption that mid-range values reflected high degrees of ambiguity while more extreme values reflected lesser ambiguity. However, based on the observed pattern of results in Study 1, this assumption may have been misguided, such that the low-similarity pair was actually more ambiguous than the high-similarity pair. That is to say, lower baseline similarity ratings and match judgment percentages may not equate to lesser ambiguity. Future research can shed light on this issue by testing a wider range of different handwriting sample pairs with respect to their pre-rated similarity. This will clarify the relationship between similarity and ambiguity, as well as provide a more nuanced understanding of the moderating effect of similarity.

It is also possible that similarity ratings and match judgments were not independent in Study 1. Instead, participants may have calibrated them to be consistent with each other, such that either of these judgments could have impacted the other. For example, those who received the high-similarity pair gave higher similarity ratings, which may have in turn increased the likelihood of judging the two samples as a match – akin to an anchoring effect (Tversky & Kahneman, 1974). Likewise, participants who were aware of the confession were more likely to judge the samples as a match, which may have encouraged them to give higher similarity ratings so as to reinforce this judgment. In contrast, those who were given a denial statement and two
ostensibly dissimilar handwriting samples had little basis for giving either a high similarity rating or a definitive match judgment. Because similarity ratings and match judgments were given at the same time, the data at hand do not allow me to determine if either of these measures impacted the other, and therefore this is purely speculative. Future research should vary the order in which discrete judgments are made in order to explore whether these judgments are independent of each other. Such an order effect would be consistent with others who have argued that forensic analyses should proceed in a “linear” rather than “circular” fashion (Dror, 2009).

Although I predicted that revisiting the case summary would exacerbate bias, I found only limited support for this hypothesis. Knowledge of the confession affected both definitive match judgments and match-confidence composite scores at Time 2, but not at Time 1. However, observed changes in similarity ratings, match judgments, and composite scores over time were minimal. I have identified three non-mutually exclusive explanations for this low rate of change. First, the feedback given to participants between their first and second round of handwriting judgments was quite benign, as participants were merely asked to review the case summary again and to provide a final set of judgments. In real-world investigations, police often overtly communicate to forensic examiners that an inculpatory judgment is desired (e.g., Risinger et al., 2002). The feedback in Study 1 may have failed to simulate real-world investigator-examiner communications insofar as it did not explicitly state a preference for any particular outcome. As such, a more suggestive communication may have produced the hypothesized effect.

Second, the feedback manipulation may have been weak because it was given in writing. Classic research indicates that obedience to authority and other forms of social impact are more likely to occur as the strength and immediacy (e.g., physical proximity) of the agent of influence increases (Latané, 1981; Milgram, 1974). Therefore, the effect of the feedback may have been
greater if the feedback had been delivered to participants in person by the experimenter. Finally, the observed lack of change in handwriting judgments over time may be partly attributable to an anchoring effect (Tversky & Kahneman, 1974). That is to say, once participants produced an initial set of judgments, they may have later been reluctant to stray from these values merely by virtue of having already produced these same judgments once previously.

Lastly, dispositional Need for Cognition (NFC) did not moderate the biasing effect of the confession on handwriting judgments. This stands in contrast to Kassin et al. (1990), who found that high-NFC individuals were more heavily persuaded than their low-NFC counterparts by information that preceded evaluations of ambiguous evidence. Importantly, statistical tests of this hypothesis in Study 1 were severely underpowered (Cohen, 1990), and thus it is possible that my small sample size precluded the detection of the predicted moderation effect.

If in fact NFC did not moderate the effect of the confession on handwriting judgments in Study 1, this null effect may be attributable to self-selection, such that high-NFC individuals were over-represented in the sample. It is certainly conceivable that individuals who are higher in NFC are more likely to seek out and complete online research studies. Consistent with this possibility, only 14.15% of participants in Study 1 generated a negative NFC score, suggesting that those who were classified as “low-NFC” may not have truly been low in NFC. Furthermore, being asked to assume the role of a mock forensic examiner may have produced a situational demand for high NFC that superseded participants’ dispositional tendencies. This is not necessarily a limitation of Study 1. If one assumes that professional forensic examiners are relatively high in NFC (dispositionally and/or situationally), the use of high-NFC samples and/or the situational induction of high NFC may actually benefit external validity.
CHAPTER 8: STUDY TWO METHOD

Study 2 tested how the use of an evidence lineup (in which a sample is compared against several comparison samples; Wells et al., 2013) affected judgments of handwriting evidence relative to the standard method of presentation (in which a sample is compared against only one comparison sample). Some researchers have argued that the latter procedure – which is akin to an eyewitness “showup” (Dysart & Lindsay, 2007) – inherently invites match judgments (e.g., Whitman & Koppl, 2010) and thus, like an eyewitness showup, increases the risk of an innocent person being misidentified as guilty (Steblay et al., 2003). As such, evidence lineups may yield benefits similar to eyewitness lineups in terms of reducing systematic error, and may also reduce the impact of biasing case information on examiners’ judgments (Risinger et al., 2002).

Participants in Study 2 served as mock handwriting examiners and evaluated handwriting samples as part of a bank robbery investigation. They first read a case summary in which either the suspect confessed to the crime (suggesting guilt; e.g., Kassin & Neumann, 1997), an alibi corroborated the suspect’s denial (suggesting innocence; see Olson & Wells, 2004), or the suspect merely denied involvement (suggesting neither guilt nor innocence).

Then, they analyzed handwriting evidence presented in one of two ways. Some viewed a showup of two handwriting samples – one from the suspect, and one robbery note written by the perpetrator – and indicated whether or not they believed that the suspect had written the robbery note. Others viewed a lineup of four samples – one from the suspect, and three robbery notes (including one unspecified note that was written by the perpetrator) – and indicated which, if any, of the three robbery notes they believed had been written by the suspect.

For some participants, the showup or lineup included one robbery note that had in fact been authored by the suspect (i.e., a “target” sample was present); for others, none of the robbery note(s) that they were given had been authored by the suspect (i.e., no “target” sample was
present). Study 2 thus employed a 3 (Context: Alibi, Denial, or Confession) X 2 (Presentation: Showup vs. Lineup) X 2 (Target: Present vs. Absent) between-subjects design.

**Pilot Study**

In Study 1, participants’ handwriting judgments were very “conservative”: Even the “high similarity” samples elicited an overall similarity rating below the midpoint of the scale and were judged as a match by only one in five participants. Accordingly, I wondered if people had unrealistic expectations: If they under-appreciated the fact that the same person’s handwriting naturally varies over time, they may have over-valued dissimilarities between two samples and thus been hesitant to deem them a match if they were less than perfectly identical.

Because a primary aim of Study 2 was to test the effects of inculpatory and exculpatory case information on match judgments, the low frequency of match judgments in Study 1 raised concern over floor effects. In an effort to overcome participants’ high response threshold for making match judgments, Study 2 was preceded by a pilot study in which we modified the task instructions to remind participants that “no one’s handwriting is always the same from one occasion to another” and therefore “even two samples written by the same person will not be perfectly identical” (see Appendix D). Participants (N = 44) were randomly assigned to one of two groups (Confession: Present vs. Absent) and were shown the high-similarity pair of samples used in Study 1 (see Appendix F). The procedure was otherwise identical to that of Study 1.

As a result of the modified instruction, participants in the pilot study produced an overall mean similarity rating of 6.05 (SD = 2.46), which was higher than the corresponding mean from Study 1 (M = 4.92, SD = 2.83), t(196) = 2.96, p = .004, d = 0.43 [95% CI: 0.06 – 0.80]. In terms of trichotomous match judgments, a one-way chi-square test using the category frequencies from Study 1 as expected frequencies showed that the distribution of match judgments was different.
from Study 1, $\chi^2(2) = 26.72$, $p < .001$, $\phi = .55$. Though the rate of CBD judgments remained constant (15.90% vs. 12.27%), match judgments were made more often (39.77% vs. 19.54%) and non-match judgments were made less often (44.32% vs. 68.18%) in the pilot study. The modified instructions were thus effective in lowering participants’ response threshold, and would therefore allow for the detection of both increases and decreases in match judgments as a function of exposure to inculpatory and exculpatory case information (respectively) in Study 2.

**Participants and Design**

A sample of 473 individuals was obtained via Amazon Mechanical Turk (mTurk; see Mason & Suri, 2012). As in Study 1, participation was restricted to U.S residents only. Each participant was randomly assigned to one of 12 cells produced by the 3 (Context: Alibi, Denial, or Confession) X 2 (Presentation Format: Showup vs. Lineup) X 2 (Target: Present vs. Absent) completely between-subjects design.

Twenty participants (4.23%) were later excluded after it was determined that they had previously participated in either Study 1 or the pilot study, and would thus have prior knowledge of the materials and aims of Study 2. I excluded an additional four participants (0.88%) who indicated that they had completed Study 2 on a cellular phone, as I was concerned that this did not give an adequate display of the handwriting stimuli. Lastly, I excluded a total of 59 participants (13.14%) who responded incorrectly to manipulation check questions which asked whether or not the suspect had confessed (3.12%) and/or whether anyone was able to vouch for the suspect’s whereabouts during the crime (10.47%). Study 2 was thus based on a final sample of $N = 390$, with individual cell sizes ranging from $n = 29$ to $n = 35$.

The final sample had a mean age of 35.04 ($SD = 12.09$; Range = 18 – 74) and a roughly equal gender distribution (47.18% female). Four participants (1.03%) reported that they are not
currently U.S. citizens. However, all were current U.S. residents, as this was a prerequisite for participation, and the sample included at least one resident from 43 of the 50 U.S. states. With respect to race, most self-identified as White (78.46%), with relatively fewer identifying as Asian (8.21%), Black (6.15%), Hispanic (3.59%), multi-racial (2.56%), and Native American (0.51%). In terms of educational attainment, 25.90% did not hold a college degree, 22.05% held a two-year college degree, 38.46% held a four-year college degree, and 13.59% held a graduate degree.

**Procedure**

As in Study 1, participants completed Study 2 using an online survey website, and were compensated for their participation with a $0.50 credit to their mTurk account. After providing informed consent (see Appendix A), participants answered basic demographic questions in which they reported their age, gender, race, U.S. citizenship, state of residence, and educational attainment. Participants were also asked to report the type of device on which they were completing the study, in order to ensure that participants were utilizing devices that provided an adequate display of the handwriting stimuli (i.e., not cellular phones; see Appendix C).

Participants read the modified instructions that were used in the pilot study (see Appendix D). These explained that they would assume the role of a handwriting identification expert whose job is to offer opinions as to whether or not handwriting samples were authored by the same person. During this time, they were reminded that handwriting varies over time such that “even two samples written by the same person will not be perfectly identical.”

Next, participants read a simulated police memo that requested their assistance with an ongoing armed robbery investigation, which was followed by one of three summaries of a bank robbery investigation (see Appendix E). The three versions of the case summary were identical to each other – and identical to the case summary used in Study 1 – with the exception of the
final paragraph which described the outcome of Hines’ police interview (i.e., the Context manipulation). By random assignment, one group (confession condition) was told that Hines confessed to the robbery but later recanted his confession. A second group (denial condition) was told that Hines maintained his innocence and claimed that he was alone at a local restaurant during the robbery. A third group (alibi condition) was told that police interviewed a cashier who confirmed that Hines was present in the restaurant on the morning of the robbery.

After reading the case summary but prior to receiving handwriting evidence, participants read instructions that were modeled after best practice guidelines for eyewitness showups and lineups (Technical Working Group on Eyewitness Evidence, 1999; Wells, 2006; see Appendix I). By random assignment, they then compared Hines’ Miranda waiver either against the note used in the robbery (showup condition) or against three robbery notes that were written by different authors (lineup condition). Also by random assignment, participants either were (target-present condition) or were not (target-absent condition) shown one robbery note that had been written by the same author as the Miranda waiver (i.e., was a “match”).

To be exact, participants in the target-present showup condition compared the Miranda waiver against the note used in the robbery which, unknown to them, was written by the same author as the Miranda waiver. Participants in the target-absent showup condition compared the waiver against the note used in the robbery, which was written by someone other than the author of the waiver. Participants in the target-present lineup condition compared the waiver against three robbery notes, one of which was written by the same author as the waiver (and two were not). Participants in the target-absent lineup condition compared the waiver against three robbery notes, none of which was written by the same author as the waiver (see Appendix J).
Participants were given an unlimited amount of time to examine the samples, and were asked to provide two judgments. First, those in the showup conditions reported whether they believed that Hines had written the lone robbery note, while those in the lineup conditions were asked which – if any – of the three robbery notes they believed was written by Hines. Second, participants in all conditions reported their confidence in this judgment. In addition, participants were given the chance to explain their rationale for these judgments in an open-ended fashion.

As in Study 1, participants also answered two self-report items regarding the extent to which their judgments were influenced by the handwriting samples and by the facts of the case, and completed a multiple-choice comprehension test to ensure that they read, understood, and recalled the details of the case summary (see Appendix H). All were then fully debriefed.

Materials

Case summary and context manipulation. Two of the three versions of the case summary that were used in Study 2 were identical to those used in Study 1: Participants in the confession and denial conditions received the confession-present and confession-absent case summaries from Study 1, respectively. Participants in the alibi condition received a third version of the case summary, which was identical to the other two versions with the exception of the final paragraph describing the outcome of the police interview of Hines (see Appendix E).

Like participants in the denial condition, those in the alibi condition were told that Hines maintained his innocence throughout the interview and claimed to have been alone at a nearby restaurant when the robbery occurred. They were also told that police interviewed the cashier who was working in the restaurant at the time of the robbery. When shown a photo of Hines, the cashier said that she remembered seeing Hines because he was the first customer during her shift that day. To ensure that the level of detail was comparable between conditions, the cashier in the
alibi condition recounted the same details that were presented by Hines in the denial condition, including what he ate, what he did while he was there, how long he stayed in the restaurant, and the fact that he was alone. According to the taxonomy of alibi strength described by Olson and Wells (2006), the cashier’s alibi would be considered moderately believable, insofar as she was not motivated to lie on Hines’ behalf, but she did not produce physical evidence to support her alibi statement and her memory could be mistaken due to her lack of familiarity with Hines.

Pre-showup/lineup instructions. Before receiving the handwriting samples, participants were instructed in a manner that was modeled after common best practice recommendations for eyewitness identification procedures (e.g., Smalarz & Wells, 2012; Wells, 2006; Wells et al., 2006; see Appendix I). These instructions explained the nature of the handwriting comparison task and included four important caveats that were adapted from the U.S. Department of Justice’s (DoJ; Technical Working Group on Eyewitness Evidence, 1999) official guidelines.

First, participants in the showup conditions were reminded that the robbery note may or may not have been written by Hines. Likewise, those in the lineup conditions were reminded of the possibility that none of the three robbery notes was written by Hines. Adding an explicit warning to this effect has been shown to reduce the risk of mistaken eyewitness identifications (Malpass & Devine, 1981; Steblay, 1997) and is now widely viewed as a beneficial practice (Wells, Steblay, & Dysart, 2012). Second, participants were reminded that “it is just as important to clear innocent persons from suspicion as it is to identify the guilty party.” Third, participants were reminded that a person’s handwriting “may not appear exactly the same on two different occasions.” This was done to mimic the practice of reminding eyewitnesses that the appearance of individuals in a lineup may have changed since the time of the incident in question. Though some have questioned the value of an explicit “appearance-change” instruction (e.g., Charman &
Wells, 2007), it is nonetheless included among the Department of Justice guidelines. Fourth, participants were assured that police would continue to investigate the incident “regardless of whether or not [they] are able to give a definitive opinion” of the handwriting evidence.

**Handwriting stimuli.** From pilot testing by Kukucka and Kassin (in press), I identified four robbery notes that did not differ in terms of their perceived similarity to the *Miranda* waiver sample ($p = .69$). These four robbery notes included one note that was written by the same author as the *Miranda* waiver (i.e., the target sample [T]) and three notes that were authored by other individuals (i.e., three “fillers,” which were arbitrarily designated as fillers A, B, and C).

Participants in the lineup conditions viewed Hines’ *Miranda* waiver with three robbery notes in a single horizontal row beneath it. This was done to ensure that all three robbery notes were equidistant from the *Miranda* waiver and to ensure that participants could view all four handwriting samples on their screen at once. In the *target-present lineup* condition, participants viewed the *Miranda* waiver along with the target sample (T) and two of the three filler samples (A, B, and/or C). At random, participants were shown one of three arrangements of the target and two filler samples, namely (from left to right), TCB, CTA, or ABT. Similarly, in the *target-absent lineup* condition, participants viewed the *Miranda* waiver along with one of three arrangements of the three filler samples, namely, ACB, BAC, or CBA (see Appendix J).

This type of randomization had several advantages. Advocates of stimulus sampling (e.g., Wells & Windschitl, 1999) have argued that using multiple exemplars of a stimulus category (i.e., lineups) enhances both external and construct validity. Also, some research has found that the mere position of an individual’s photograph in an eyewitness lineup can affect its likelihood of being identified as the perpetrator (e.g., Sporer, 1993). Similarly, some studies have found eyewitness identifications to be influenced by which filler photos are shown adjacent to the
target (e.g., Gonzalez, Davis, & Ellsworth, 1995). By varying the lineup positions of the target and filler samples within each Target condition, I controlled for the possibility that some positions were inherently more appealing than others, and ensured that the target sample would appear next to each of the filler samples in at least one of the three target-present arrangements.

Participants in both showup conditions were shown the Miranda waiver with a single robbery note beneath it. This lone robbery note was shown in the same size as when it was presented as part of a lineup. In the target-present showup condition, all participants viewed the Miranda waiver and the target robbery note. In the target-absent showup condition, they viewed the Miranda waiver and one of the three filler notes (A, B, or C), selected at random. By varying which of the three filler samples was paired with the waiver in target-absent showups, I again used stimulus sampling as a means of enhancing validity (Wells & Windschitl, 1999).

**Dependent measures.** Participants first made a discrete judgment as to the authorship of the handwriting samples. For Showups, participants were asked whether or not they believed that Mr. Hines had written the note used in the robbery, with response options of “yes,” “no,” and “cannot be determined.” For Lineups, participants were asked which, if any, of the three robbery notes they believed to be written by Mr. Hines, with response options for each of the three notes, as well as options of “none of these” and “cannot be determined.”

Participants rated their confidence in this judgment, using a scale from 1 (not at all confident) to 10 (very confident). Participants who concluded that the presence of a matching sample “cannot be determined” were assigned a confidence rating of zero. Participants were then given an opportunity to explain the rationale for their judgment in an open-ended fashion.

**Comprehension test.** Participants answered seven multiple-choice questions to ensure that they read, understood, and recalled the content of the case summary. These included the five
questions from Study 1, along with two new items which asked whether the suspect appeared to know the details of the crime and whether anyone could vouch for Hines’ whereabouts when the crime occurred (see Appendix H). The former aimed to measure the extent to which participants appreciated the content of the confession, thereby supplementing the critical item which merely asked whether participants recalled the confession (or lack thereof). The latter aimed to measure participants’ awareness of the presence or absence of alibi evidence.

A total of 59 participants (13.14%) were excluded after failing to accurately recall the presence or absence (depending on condition) of a confession (3.12%) and/or an alibi statement (10.47%). In light of the large proportion of participants who were excluded on the basis of these items, we chose to retain 27 additional participants (6.92%) who correctly recalled that Hines did or did not confess, but inaccurately reported that Hines either did or did not appear to know the details of the crime. Of these, 21 reported that Hines had confessed but did not appear to know the details of the crime, and six (four in the Denial condition; two in the Alibi condition) reported that Hines did not confess but did appear to know the details of the crime.

Hypotheses

**H1:** Consistent with the meta-analytic comparison of eyewitness showups and lineups by Steblay et al. (2003), I predicted that Presentation would affect choosing rates (i.e., the proportion of participants who identify any of the samples as a match, regardless of accuracy), such that participants would be more likely to choose from a Lineup than from a Showup, and that this difference would be found both when the Target was present and when it was absent.

**H2:** Given that confession evidence implies the suspect’s guilt (e.g., Kassin et al., 2013) and alibi evidence implies the suspect’s innocence (e.g., Olson & Wells, 2004), I predicted that Context would affect choosing rates, such that participants in the Confession condition would
show the highest choosing rates, and those in the Alibi condition the lowest choosing rates, both overall and within each level of Presentation (i.e., Showups and Lineups).

**H3:** In line with Steblay et al. (2003), I predicted that Showups would produce a higher overall proportion of accurate judgments than Lineups, but this effect would be moderated by Target. When the Target was absent, I predicted that participants who used a Lineup would be more likely to misidentify a non-matching sample as a match than those who used a Showup. (Notably, the findings of Steblay et al. imply the opposite of Miller [1987], who found that evidence lineups decreased the likelihood of identifying a non-matching sample as a match.) When the Target was present, I predicted that participants who used a Showup would be more likely to misreport that the Target was absent than those who used a Lineup.

**H4:** Supporting the argument that the use of an evidence lineup will protect examiners against the influence of biasing case information (e.g., Kassin et al., 2013; Risinger et al., 2002), I predicted that the effect of Context on judgment accuracy would be moderated by Presentation and Target. For Target-Present Showups, I predicted that the frequency of accurate judgments would be highest when participants were told of the Confession, and lowest when they were told of the Alibi. For Target-Absent Showups, I predicted that this pattern would be reversed. For Lineups, I did not predict an effect of Context on the frequency of accurate judgments.

**H5:** Consistent with research on the relationship between the accuracy and confidence of eyewitness lineup identifications (e.g., Sporer, Penrod, Read, & Cutler, 1995), I predicted a weak positive correlation between the accuracy and confidence of match judgments from Lineups.

**H6:** I predicted that Study 2 would replicate the “illusion of objectivity” effect found in Study 1 (Kunda, 1990), such that participants who used Showups would rate their handwriting judgments as being influenced more by handwriting itself than by the facts of the case.
CHAPTER 9: STUDY TWO RESULTS

The frequencies of all handwriting judgments across all conditions can be found in Table 1. Overall, 341 participants (87.44%) gave a definitive judgment of the handwriting evidence, while the remaining 49 participants (12.56%) gave judgments of “cannot be determined” (CBD). The frequency of CBD judgments did not differ as a function of Context, $\chi^2(2) = 4.19, p = .123, \phi = .10$, Presentation, $\chi^2(1) = 2.00, p = .158, \phi = .07$, or Target, $\chi^2(1) = 1.14, p = .285, \phi = .05$.

For Showups, Context had no effect on the overall frequency of CBD judgments, $\chi^2(2) = 3.87, p = .144, \phi = .14$, nor did it have an effect when the target was present, $\chi^2(2) = 0.95, p = .621, \phi = .10$, or absent, $\chi^2(2) = 4.17, p = .124, \phi = .21$. Similarly for Lineups, Context had no effect on the overall frequency of CBD judgments, $\chi^2(2) = 2.14, p = .343, \phi = .11$, nor did it have an effect when the target was present, $\chi^2(2) = 0.58, p = .748, \phi = .08$, but it had a marginal effect when the target was absent, $\chi^2(2) = 5.56, p = .062, \phi = .24$, such that CBD judgments were more common in the Confession condition (22.86%) than in the Denial condition (3.33%).

Because CBD judgments appeared uniformly distributed among the twelve cells in the 3 (Context) X 2 (Presentation) X 2 (Target) design, and because the primary focus of Study 2 was on the relative frequency and confidence of accurate and inaccurate judgments (which CBD cannot rightly be considered as either), I excluded those 49 participants who produced CBD judgments from all subsequent analyses, unless otherwise noted.

Choosing Rates

Choosing was defined as identifying any one of the robbery notes as a match. Choosing is thus independent of accuracy, as a choice can be accurate (i.e., identifying the target) or inaccurate (i.e., identifying a filler sample); the absence of a choice can also be accurate (i.e., when the target was not present) or inaccurate (i.e., when the target was present). Observed
choosing rates for each of the 12 cells can be found in Table 2. Across all conditions, 43.59% of participants were choosers (with CBD judgments classified as non-choosers).

Supporting Hypothesis 1, Presentation impacted overall choosing rates, such that Lineups (57.14%) produced more choosing than did Showups (29.90%), $\chi^2(1) = 29.43, p < .001, \phi = .28, OR = 3.13 [95% CI: 2.06 – 4.75]$. As predicted, Lineups produced more choosing than Showups both when the Target was absent, $\chi^2(1) = 24.20, p < .001, \phi = .35, OR = 4.51 [95% CI: 2.43 – 8.34]$, and when it was present, $\chi^2(1) = 7.76, p = .005, \phi = .20, OR = 2.25 [95% CI: 1.27 – 4.01]$. Furthermore, Lineups produced more choosing than Showups in the Alibi, $\chi^2(1) = 9.02, p = .003, \phi = .26, OR = 3.02 [95% CI: 1.45 – 6.30]$, and Denial, $\chi^2(1) = 22.50, p < .001, \phi = .43, OR = 7.07 [95% CI: 3.02 – 16.57]$, conditions, and marginally more choosing in the Confession condition, $\chi^2(1) = 3.48, p = .062, \phi = .16, OR = 1.90 [95% CI: 0.97 – 3.74]$. In partial support of Hypothesis 2, Context had an overall effect on choosing, $\chi^2(2) = 9.03, p = .011, \text{Cramér’s } V = .15$, such that the Confession (53.62%) produced more choosing than did the Alibi (39.69%), $\chi^2(1) = 5.24, p = .022, \phi = .14, OR = 1.76 [95% CI: 1.08 – 2.85]$, or Denial (36.36%), $\chi^2(1) = 7.74, p = .005, \phi = .17, OR = 2.02 [95% CI: 1.23 – 3.33]$ -- which did not differ from each other, $\chi^2(1) = 0.30, p = .586, \phi = .03, OR = 1.15 [95% CI: 0.69 – 1.92]$. However, the effect of Context on choosing rates was moderated by both Presentation and Target. Context did not affect choosing for Lineups, $\chi^2(2) = 1.19, p = .552, \text{Cramér’s } V = .08$, but did affect choosing for Showups, $\chi^2(2) = 13.94, p = .001, \text{Cramér’s } V = .27$, such that choosing occurred more often in the presence of a Confession (45.59%) than in the presence of an Alibi (26.56%), $\chi^2(1) = 5.16, p = .023, \phi = .20, OR = 2.32 [95% CI: 1.11 – 4.82]$, or a Denial, $\chi^2(1) = 13.04, p < .001, \phi = .32, OR = 4.36 [95% CI: 1.90 – 9.97]$. Second, Context did not affect choosing when the Target was absent, $\chi^2(2) = 3.98, p = .137, \text{Cramér’s } V = .14$, but did affect
choosing when it was present, $\chi^2(2) = 6.53, p = .042$, Cramér’s $V = .18$, such that choosing occurred more often in the presence of a Confession (55.71%) than in the presence of a Denial (33.87%), $\chi^2(1) = 6.33, p = .012$, $\varphi = .22$, $OR = 2.46$ [95% CI: 1.21 – 4.98].

Within-row comparisons were conducted on the choosing frequencies in Table 2 to explore the possible interaction of all three independent variables with respect to choosing. For target-present showups, Confession led to more choosing than either Denial or Alibi; similarly, for target-absent showups, Confession produced more choosing than did Denial. In contrast, Context had no impact on choosing for either target-absent or target-present lineups.

**Judgment Accuracy**

To test the effects of the manipulations on response accuracy, I dichotomized each participant’s judgment as either accurate (correctly identifying the target when present; correctly rejecting the lineup/showup when the target was absent) or inaccurate (i.e., failing to identify the target when present; identifying a filler sample as the target when the target was absent).

Accuracy rates for each of the 12 cells can be found in Table 3. Across all conditions, 40.76% of participants made accurate judgments (with CBD judgments excluded).

Supporting Hypothesis 3, Showups produced better overall accuracy (55.76%) than did Lineups (26.70%), $\chi^2(1) = 29.77, p < .001$, $\varphi = .30$, $OR = 3.45$ [95% CI: 2.17 – 5.56]. Contrary to Hypothesis 3, Showups produced greater accuracy both when the Target was absent, $\chi^2(1) = 25.45, p < .001$, $\varphi = .39$, $OR = 5.24$ [95% CI: 2.70 – 10.10], and when it was present, $\chi^2(1) = 7.99, p = .005$, $\varphi = .21$, $OR = 2.58$ [95% CI: 1.32 – 5.03]. Also, Showups produced greater accuracy within all three Context conditions -- including Alibi, $\chi^2(1) = 4.77, p = .029$, $\varphi = .20$, $OR = 2.27$ [95% CI: 1.09 – 4.76], Denial, $\chi^2(1) = 9.59, p = .002$, $\varphi = .30$, $OR = 3.57$ [95% CI: 1.56 – 8.33], and Confession, $\chi^2(1) = 17.71, p < .001$, $\varphi = .39$, $OR = 5.56$ [95% CI: 2.44 – 12.50].
Context had no overall effect on accuracy, $\chi^2(2) = 0.93, p = .628$, Cramér’s $V = .05$, indicating that accuracy rates were equivalent between the Alibi (44.17%), Denial (39.62%) and Confession (38.26%) conditions. Context did not affect the overall accuracy of judgments from Showups, $\chi^2(2) = 0.20, p = .904$, Cramér’s $V = .04$, or Lineups, $\chi^2(2) = 3.28, p = .194$, Cramér’s $V = .14$, conditions. However, Target moderated the effect of Context on accuracy (see Figure 3). When the Target was absent, Context impacted accuracy, $\chi^2(2) = 10.27, p = .006$, Cramér’s $V = .25$, such that the Confession condition was less accurate than both the Alibi, $\chi^2(1) = 9.26, p = .002, \phi = .29, OR = 3.26 [95\% CI: 1.51 – 7.04]$, and Denial, $\chi^2(1) = 6.07, p = .014, \phi = .24, OR = 2.69 [95\% CI: 1.21 – 5.95]$, conditions -- which did not differ, $\chi^2(1) = 0.26, p = .610, \phi = .05, OR = 1.21 [95\% CI: 0.58 – 2.56]$. When the Target was present, Context again affected accuracy, $\chi^2(2) = 6.58, p = .037$, Cramér’s $V = .20$, such that the Confession condition was now more accurate than both the Alibi, $\chi^2(1) = 2.85, p = .050, \phi = .18, OR = 2.15 [95\% CI: 0.99 – 4.65]$, and Denial, $\chi^2(1) = 5.27, p = .022, \phi = .21, OR = 2.56 [95\% CI: 1.14 – 5.78]$, conditions -- which again did not differ, $\chi^2(1) = 0.16, p = .693, \phi = .04, OR = 1.19 [95\% CI: 0.50 – 2.85]$.

Within-row comparisons were conducted on the accuracy rates in Table 3 to test the prediction of Hypothesis 4 that Presentation, Target, and Context would interact to influence accuracy. Supporting Hypothesis 4, Confession produced a higher accuracy rate than both Denial and Alibi for target-present Showups, and produced a lower accuracy rate than Denial (neither of which differed from Alibi) for target-absent Showups. Contrary to Hypothesis 4, Confession produced a lower accuracy rate than Alibi for target-absent lineups (neither of which differed from Denial), whereas Context had no effect on accuracy for target-present lineups.
Judgment Confidence

Across all conditions, judgment confidence was marginally and weakly correlated with judgment accuracy, \( r(338) = .10, p = .056 \). Contrary to Hypothesis 5, judgment confidence did not predict judgment accuracy within either Showups, \( r(163) = .07, p = .352 \), or Lineups, \( r(174) = .10, p = .202 \). Confidence likewise did not predict accuracy in the Alibi, \( r(117) = .12, p = .185 \), Denial, \( r(104) = .06, p = .542 \), or Confession, \( r(113) = .14, p = .145 \), conditions, nor did it predict accuracy among participants who viewed target-absent showups, \( r(78) = .02, p = .849 \), target-present showups, \( r(83) = .14, p = .213 \), target-absent lineups, \( r(84) = .19, p = .085 \), or target-present lineups, \( r(87) = .02, p = .885 \).

A 3 (Context) X 2 (Presentation) X 2 (Target) ANOVA was performed on confidence ratings irrespective of accuracy to explore any overall effects of the manipulations on judgment confidence. A main effect of Context was found, \( F(2,328) = 3.80, p = .023, \eta_p^2 = .02 \). Post hoc Tukey analyses showed that participants in the Denial condition were significantly less confident (\( M = 6.29, SD = 2.17 \)) than those in the Alibi (\( M = 6.90, SD = 1.98 \)) and Confession (\( M = 6.95, SD = 1.62 \)) conditions, which did not differ. Presentation had no effect on confidence, \( F(1,328) = 2.02, p = .157, d = 0.16 \) [95% CI: -0.05 – 0.36], as confidence ratings did not differ between Showups (\( M = 6.88, SD = 1.95 \)) and Lineups (\( M = 6.58, SD = 1.94 \)). No effect of Target was found, \( F(1,328) = 1.26, p = .262, d = 0.13 \) [95% CI: -0.08 – 0.34], as confidence ratings were equal when the target was absent (\( M = 6.60, SD = 2.00 \)) and present (\( M = 6.85, SD = 1.89 \)). None of the two-way interactions was significant – including Presentation X Target, \( F(1,328) = 0.69, p = .408, \eta_p^2 = .00 \), Presentation X Context, \( F(2,328) = 0.59, p = .557, \eta_p^2 = .00 \), and Target X Context, \( F(2,328) = 0.73, p = .484, \eta_p^2 = .00 \) – nor was the three-way interaction significant.
$F(2,328) = 0.94, p = .391, \eta^2_p = .01$. This pattern of results remained identical regardless of whether or not CBD judgments (for which confidence = 0) were included.

**Accuracy-Confidence Composite**

To obtain a more sensitive test of the effects of the manipulations on judgment accuracy, I generated an accuracy-confidence composite score for all participants by computing the product of their dichotomized response accuracy scores (with accurate judgments coded as +1, inaccurate judgments coded as -1, and CBD judgments excluded) and their associated continuous confidence rating (1-10), thus producing a score that could range from -10 (highly confident inaccurate response) to +10 (highly confident accurate response).

A 3 (Context) X 2 (Presentation) X 2 (Target) ANOVA on these composite scores revealed a main effect of Presentation, $F(1,377) = 32.60, p < .001, d = 0.55$ [95% CI: -0.11 – 1.21], such that composite scores were higher for Showups ($M = 0.79, SD = 6.57$) than for Lineups ($M = -2.65, SD = 5.94$). No effect of Context was found, $F(2,377) = 0.26, p = .769, \eta^2_p = .00$, such that scores were equal between the Alibi ($M = -0.58, SD = 6.87$), Denial ($M = -1.03, SD = 6.17$), and Confession conditions ($M = -1.18, SD = 6.43$). A main effect of Target emerged, $F(1,377) = 19.15, p < .001, d = 0.39$ [95% CI: -0.28 – 1.07], such that scores were higher when the target was absent ($M = 0.32, SD = 6.38$) than present ($M = -2.18, SD = 6.36$).

The main effect of Target was qualified by a significant Context X Target interaction, $F(2,377) = 9.81, p < .001, \eta^2_p = .05$ (see Figure 4). Simple effects tests indicated that accuracy-confidence scores were higher in both the Alibi, $t(128) = 4.29, p < .001, d = 0.76$ [95% CI: -0.38 – 1.90], and Denial, $t(119) = 3.86, p < .001, d = 0.71$ [95% CI: -0.39 – 1.81], conditions when the target was absent ($Ms = 1.78$ and $1.07$, $SDs = 6.48$ and $6.19$, respectively) rather than present ($Ms = -3.08$ and -3.03, $SDs = 6.41$ and $5.48$). In the Confession condition, accuracy-confidence scores...
did not differ when the target was absent ($M = -1.75$, $SD = 5.98$) versus present ($M = -0.63$, $SD = 6.83$), $t(136) = -1.03, p = .307, d = 0.18$ [95% CI: -1.00 – 1.35].

Neither the Context X Method interaction, $F(2,377) = 1.28, p = .280, \eta^2_p = .01$, nor the Method X Target interaction, $F(1,377) = 1.68, p = .196, \eta^2_p = .00$, nor the three-way interaction, $F(2,377) = 1.47, p = .232, \eta^2_p = .01$, achieved significance. (Again, this pattern of results remained identical regardless of whether CBD judgments were included.)

**Signal Detection Analysis**

Some researchers have recently advocated for the adoption of a signal detection theory (SDT) framework in eyewitness psychology research, proposing that receiver operating characteristic (ROC) analysis should be used to interpret the outcomes of lineup identification procedures (e.g., Gronlund, Wixted, & Mickes, 2014). Supporters of this approach have argued that eyewitness lineups present a signal-detection problem, insofar as the culprit (i.e., the signal/target) either is or is not present in the lineup, and eyewitnesses must make a judgment as to whether or not the culprit is present (Mickes, Flowe, & Wixted, 2012).

Although I did not perform a full ROC analysis, I analyzed the data using an SDT approach to further explore the effects of Presentation and Context on handwriting judgments. The SDT framework classifies responses into one of four categories and calculates the rates of these four response types. When the target (in this case, a matching handwriting sample) was present, responses were coded as either a “hit” (i.e., when the participant correctly identified the target) or a “miss” (i.e., when the participant failed to identify the target). The “hit rate” is the proportion of target-present trials in which the target was correctly identified. When the target was absent, responses were coded as either a “correct rejection” (i.e., when the participant correctly reported that the target was not present) or a “false alarm” (i.e., when the participant
incorrectly identified the target as present). The “false alarm rate” is the proportion of target-absent trials in which the target was incorrectly judged as being present.

The SDT framework also calculates two parameters: a sensitivity index ($d'$) and a bias index ($C$). The former represents the standardized difference between the hit rate and false alarm rate, and thus provides a measure of the readiness with which the target was detected. Larger $d'$ values indicate greater ability to discriminate between trials when the target was present and trials when it was not. The latter provides a measure of response bias, with $C < 0$ suggesting a liberal response criterion (i.e., a tendency to report that the target was present, regardless of its true presence or absence) and $C > 0$ suggesting a conservative response criterion (i.e., a tendency to report that the target was not present, regardless of its true presence or absence).

Table 4 provides the observed rates of four possible signal detection outcomes and the corresponding $d'$ and $C$ parameters, broken down by Presentation and Context. Underscoring the observed effect of Presentation on accuracy, discrimination ability was superior for Showups ($d' = 0.34$) than for Lineups ($d' = -1.26$). The negative value of $d'$ for Lineups reflects the fact that they more often produced false alarms than hits, whereas the opposite was true for Showups. Context appeared to have little effect on discrimination ability for Showups, but for Lineups, $d'$ decreased from the Alibi ($d' = -0.88$) to Denial ($d' = -1.39$) to Confession ($d' = -1.74$) conditions. This decrease appears due to an increase in false alarms from the Alibi (.52) to Denial (.69) to Confession (.85) conditions, rather than to an effect of Context on hit rate. Comparison of the $C$ values shows that judgments from both Showups ($C = 0.39$) and Lineups ($C = 0.17$) were rather conservative. However, consistent with the observed effect of Context on choosing rates, the Confession condition showed more liberal responding (i.e., participants were more likely to
conclude that the target was present, regardless of its true presence or absence) for both Showups and Lineups ($Cs = -0.15$ and $-0.17$, respectively), relative to the Alibi and Denial conditions.

**Diagnosticity**

Among eyewitness researchers, another common approach is to compute *diagnosticity ratios*, which are ratios of correct identifications from target-present lineups [i.e., $p$(Correct ID | TP)] to false identifications from target-absent lineups [i.e., $p$(False ID | TA); see Wells & Lindsay, 1980]. A diagnosticity ratio of 1.0 indicates that targets and fillers are equally likely to be identified as the perpetrator. As the diagnosticity ratio increases, so too does the informativeness of the identification with respect to the suspect’s actual guilt. As a more interpretable alternative, Dupuis and Lindsay (2007) recommended reporting the “percentage guilty,” which is the proportion of individuals identified as guilty who are in fact guilty, i.e., $p$(Correct ID | TP) / [$p$(Correct ID | TP) + $p$(False ID | TA)]. Table 5 provides $p$(Correct ID | TP), $p$(False ID | TA), diagnosticity ratio, and percentage guilty for showups and lineups both overall and within each of the three Context conditions. Unlike the preceding SDT analysis, CBD judgments were factored into the computation of the conditional probabilities in Table 5.

Wells and Olson (2002) provided a detailed account of how diagnosticity ratios can be used to model information gain from eyewitness identifications. In so doing, they described a method for testing whether a given diagnosticity ratio is different from 1.0, which relies on arcsine transformations of the conditional probabilities $p$(Correct ID | TP) and $p$(False ID | TA) to produce a $Z$ statistic whose associated $p$ value indicates whether the observed diagnosticity ratio differs significantly from 1.0 (and thus shows whether the amount of information gained from the identification procedure is significantly different from zero).
Using Wells and Olson’s (2002) method, Table 5 presents Z scores and corresponding p values for showups and lineups, both overall and within each Context condition. For showups, 60.34% of choosers produced accurate judgments, though diagnosticity was not significantly different from 1.0 ($dx = 1.46; Z = 1.70, p = .089$). For lineups, only 24.36% of choosers made accurate judgments, indicating that participants were approximately three times as likely to misidentify a filler sample as the target as they were to correctly identify the target. As such, the diagnosticity ratio for lineups ($dx = 0.34$) was significantly less than 1.0 ($Z = -5.81, p < .001$).

I also used Wells and Olson’s (2002) method to compare the diagnosticity of showups and lineups. Consistent with the observed effect of Presentation on accuracy, Showups produced greater diagnosticity than Lineups, $Z = 5.30, p < .001$. As with accuracy, diagnosticity was higher for Showups than Lineups in all three Context conditions, including Alibi, $Z = 1.99, p = .046$, Denial, $Z = 3.73, p < .001$, and Confession, $Z = 3.69, p < .001$.

**Open-Ended Responses**

Participants gave open-ended explanations for their handwriting judgments, which were coded for several features. First, I counted the number of words in each participant’s response. On average (including CBD judgments), responses were 30.80 words long ($SD = 28.92; Med = 22.50; Range = 2 – 240; $one participant gave no explanation). The raw distribution of word counts was severely positively skewed (Skewness / $SE_{skew} = 25.45$; Shapiro-Wilk $W = .71, p < .001$). These data were thus log-transformed, which corrected the observed skewness (Skewness / $SE_{skew} = -1.01$; Shapiro-Wilk $W = .99, p = .072$)

A 3 (Context) X 2 (Presentation) X 2 (Target) ANOVA on log-transformed word counts revealed a main effect of Presentation, $F(1,378) = 8.50, p = .004, d = 0.30$ [95% CI: 0.22 – 0.38], such that participants who viewed a Showup ($M = 35.23, SD = 33.00$) gave longer explanations
than those who viewed a Lineup ($M = 26.42$, $SD = 23.50$). Neither a main effect of Target, $F(1,378) = 0.01, p = .913, d = 0.00 [95\% CI: -0.08 – 0.08]$, nor a main effect of Context, $F(2,378) = 0.62, p = .539, \eta^2_p = .00$, was found. A marginal Context X Target interaction was found, $F(2,378) = 2.79, p = .063, \eta^2_p = .02$, such that Target did not affect word counts in either the Denial, $t(119) = -1.12, p = .265, d = 0.21 [95\% CI: 0.06 – 0.35]$, or Confession, $t(136) = -0.77, p = .440, d = 0.13 [95\% CI: 0.01 – 0.25]$, conditions, but explanations in the Alibi condition were marginally longer when the Target was absent (raw $M = 34.76$, $SD = 34.09$) rather than present ($M = 23.49$, $SD = 16.97$), $t(129) = 1.94, p = .055, d = 0.34 [95\% CI: 0.20 – 0.48]$. Neither the Context X Presentation interaction, $F(2,378) = 0.10, p = .908, \eta^2_p = .00$, nor the Presentation X Target interaction, $F(1,378) = 0.65, p = .420, \eta^2_p = .00$, nor the three-way interaction, $F(2,377) = 1.74, p = .177, \eta^2_p = .01$, achieved significance.

I also coded these open-ended explanations for (a) whether the response mentioned any aspects of the case summary (e.g., the alibi, confession, etc.), and (b) the number of handwriting features (e.g., individual letters or combinations) that were mentioned. To check for inter-rater reliability, a random sample of 39 responses (10\% of the total) was coded by two independent coders, who exhibited a 100\% agreement rate. As a result, I coded the remaining 351 responses.

Only nine participants (2.31\%) cited case facts as justification for their handwriting judgments. Of these, eight were in the Showup condition (four in the Alibi condition, and two each in the Denial and Confession conditions). Three participants cited Hines’ alibi, one cited his confession, three cited the lack of other physical evidence that implicated him, one cited his nervousness when stopped by police, and one cited “other evidence” without being specific.

On average, participants cited 1.38 specific handwriting features in their responses ($SD = 1.79; Med = 1; Range = 0 – 11$). A 3 (Context) X 2 (Presentation) X 2 (Target) ANOVA on the
number of features cited found no effect of Context, \( F(2,378) = 1.06, p = .348, \eta^2_p = .01 \), nor an effect of Presentation, \( F(1,378) = 0.92, p = .339, d = 0.10 \) [95% CI: -0.08 – 0.28], nor an effect of Target, \( F(1,378) = 0.02, p = .893, d = 0.03 \) [95% CI: -0.15 – 0.21]. However, a significant Context X Target interaction was found, \( F(2,378) = 3.51, p = .031, \eta^2_p = .02 \) (see Figure 6). Simple effects tests indicated that in the Confession condition, participants cited more features when the Target was present \( (M = 1.63, SD = 1.70) \) than when it was absent \( (M = 1.04, SD = 1.32) \), \( t(136) = 2.25, p = .026, d = 0.39 \) [95% CI: 0.13 – 0.64]. In the Alibi condition, they cited marginally fewer features when the Target was present \( (M = 0.95, SD = 1.37) \) than when it was absent \( (M = 1.51, SD = 1.96) \), \( t(129) = 1.89, p = .061, d = 0.33 \) [95% CI: 0.04 – 0.62]. In the Denial condition, they cited the same number of features whether the Target was present \( (M = 1.61, SD = 2.44) \) or absent \( (M = 1.53, SD = 1.69) \), \( t(119) = 0.23, p = .820, d = 0.04 \) [95% CI: -0.33 – 0.41]. Neither the Context X Presentation interaction, \( F(2,378) = 1.17, p = .311, \eta^2_p = .01 \), nor the Presentation X Target interaction, \( F(1,378) = 0.32, p = .570, \eta^2_p = .00 \), nor the three-way interaction, \( F(2,378) = 1.19, p = .305, \eta^2_p = .01 \), achieved significance.

**Self-Reported Influence**

A 3 (Context) X 2 (Presentation) X 2 (Target) X 2 (Source: Handwriting vs. Case Facts) mixed ANOVA, with repeated measures on the fourth factor, was used to test the impact of the manipulations on participants’ (including those who gave CBD judgments) self-reported influence by the handwriting samples and case facts on handwriting judgments. A main effect of Source provided support for Hypothesis 6, \( F(1,378) = 195.38, p < .001, d = 0.70 \) [95% CI: 0.59 – 0.81], as participants reported being influenced more by the handwriting samples \( (M = 7.83, SD = 2.24) \) than by the case facts \( (M = 5.04, SD = 2.76) \).
This main effect of Source was qualified by a significant Context X Source interaction, $F(2,378) = 3.68, p = .026, \eta^2_p = .02$ (see Figure 5). Dependent $t$-tests for simple effects indicated that participants felt more influenced by the handwriting samples than by the case facts in the Alibi, $t(130) = 6.92, p < .001, d = 0.61$ [95% CI: 0.41 – 0.80], Denial, $t(120) = 10.39, p < .001, d = 0.94$ [95% CI: 0.74 – 1.15], and Confession, $t(137) = 7.02, p < .001, d = 0.60$ [95% CI: 0.41 – 0.79] conditions, with the magnitude of this difference being largest in the Denial condition.

Stated differently, participants in all three Context conditions reported being equally influenced by the handwriting samples, $F(2,387) = 1.65, p = .193, \eta^2_p = .01$. However, Context affected the extent to which they felt influenced by the case facts, $F(2,387) = 3.90, p = .021, \eta^2_p = .02$, with post hoc Tukey analyses indicating that they rated the case facts as more influential in the Alibi condition ($M = 5.46, SD = 2.74$) than in the Denial condition ($M = 4.50, SD = 2.76$), neither of which differed from the Confession condition ($M = 5.12, SD = 2.73$).

Source did not interact with either Presentation, $F(1,378) = 0.19, p = .659, \eta^2_p = .00$, or Target, $F(1,378) = 0.46, p = .497, \eta^2_p = .00$, and none of the three-way interactions nor the four-way interaction was significant, all $Fs < 1.80$, all $ps > .16$, all $\eta^2_p < .01$. 

87
CHAPTER 10: STUDY TWO DISCUSSION

Consistent with research on eyewitness identification procedures (Steblay et al., 2003), the use of an evidence lineup as opposed to an evidence “showup” rendered participants more likely to choose one of the samples in the lineup as a match to the suspect’s sample. In turn, evidence lineups produced lower accuracy than showups overall and within both target-absent and target-present conditions. That is to say, participants who viewed evidence lineups were less likely to correctly identify a matching sample when one was present, and were more likely to misidentify a filler sample as a match when no matching sample was present. Study 2 thus failed to replicate Miller’s (1987) finding that evidence lineups reduced the rate of false identifications.

The eyewitness psychology literature provides a possible explanation as to why evidence lineups hindered rather than benefitted judgment accuracy. It should be noted that the evidence lineups used in Study 2 were all presented in a simultaneous fashion; that is to say, participants viewed the suspect’s sample and all three robbery notes at the same time. In the first experiment to compare eyewitness identifications made from showups and simultaneous lineups, Gonzalez, Ellsworth, and Pembroke (1993) found that choosing occurred more often from simultaneous lineups relative to showups, and consequently, witnesses were more likely to misidentify a filler sample as the culprit when using a simultaneous lineup rather than a showup. To explain this finding, Gonzalez et al. argued that showups and simultaneous lineups were evoking different processing strategies: Showups elicited an absolute judgment as to whether the lone suspect is the culprit, whereas simultaneous lineups elicited a relative judgment as to which of the possible suspects is the most similar in appearance to the culprit. As a result, showups produced fewer correct and fewer false identifications relative to simultaneous lineups.
In Study 2, simultaneous evidence lineups likewise yielded consistently higher choosing rates than showups, regardless of the content of the case summary or the presence of the target sample. Simultaneous evidence lineups thus created a robust response bias such that participants became more likely to conclude that a matching sample was present, regardless of whether this was the case. Moreover, participants’ open-ended explanations for their judgments suggest that evidence lineups may have evoked relative rather than absolute processing of the handwriting samples. Despite receiving only two handwriting samples instead of four, participants who viewed evidence showups actually gave longer open-ended justifications for their judgments than those who viewed lineups. It may be that participants who viewed simultaneous lineups thus engaged in more holistic processing of the handwriting stimuli, whereas those who viewed showups made more nuanced comparisons of the samples. With this in mind, I re-examined the open-ended explanations of participants who viewed evidence lineups for indications that they used a relative rather than absolute processing strategy. Supporting this interpretation, I found that 14.29% of these participants either rationalized their judgment as the best available option (e.g., “Exhibit B is the one that looks most like Mr. Hines’ writing,” “Exhibit D is the most similar”) or described the process of arriving at a judgment as a process of elimination (e.g., “Exhibit D is the only one I didn’t immediately eliminate,” “I was unsure about Exhibit B, but the other two I eliminated”). Thus, like simultaneous eyewitness lineups, it appears that the evidence lineups used in Study 2 led many to identify the “best available” robbery note as a match to the suspect’s handwriting sample, which often led to an erroneous judgment.

The notion that simultaneous lineups can be problematic insofar as they encourage relative rather than absolute judgments can be traced back to Wells (1984). As a corrective measure, a number of eyewitness researchers have recommended that eyewitness lineups be
administered *sequentially* rather than simultaneously (e.g., Lindsay & Wells, 1985). That is to say, rather than viewing multiple suspect photos at the same time and deciding which (if any) is the culprit, a sequential lineup presents suspect photos one-at-a-time and asks witnesses to give a discrete and absolute judgment of each photo before moving on to the next. Meta-analytic comparisons of simultaneous and sequential lineups confirm that sequential presentations tend to produce identifications that are more diagnostic of guilt than those made from simultaneous presentations (Steblay, Dysart, Fulero, & Lindsay, 2001; Steblay, Dysart, & Wells, 2011). Given that simultaneous lineups and showups impacted judgments of handwriting evidence in Study 2 in the same way as they affect eyewitness identifications, evidence lineups may likewise produce more diagnostic judgments when presented sequentially rather than simultaneously.

Knowledge of a confession produced the hypothesized effect on handwriting judgments. Participants who were told that the suspect confessed had the highest overall choosing rates, and consequently showed the lowest accuracy for target-absent showups and the highest accuracy for target-present showups. This finding underscores a key argument in favor of shielding examiners from case information, namely that exposure to case information undermines the independence (and, therefore, the probative value) of the examiner’s judgment (e.g., Dror et al., 2013). These participants frequently gave accurate judgments of target-present showups not because they were skilled at comparing handwriting samples, but because their knowledge of the confession created an inference of guilt that just so happened to be correct. This phenomenon has dangerous real-world implications: Although the forensic evidence appears to independently verify the suspect’s guilt, this is merely an incidental product of the confession (see Kassin, 2012).

Although knowledge of a confession did not impact choosing rates from target-absent lineups, it did diminish judgment accuracy, such that participants in the Confession and Denial
conditions were equally likely to identify a filler sample as a match, but those in the Confession condition were less likely to correctly reject the lineup and (marginally) more likely to give a judgment of “cannot be determined.” It is possible that these participants correctly ascertained that no matching sample was present, but were reluctant to draw this conclusion due to their knowledge of the confession, which implied that a matching sample was present. So, instead of rejecting the lineup -- which would contradict the confession -- they chose not to give a definitive judgment, thereby neither affirming nor contradicting the confession.

In contrast, knowledge of alibi evidence did not have the predicted effects on handwriting judgments. This suggests that knowledge of the suspect’s alibi did not create a parallel belief in the suspect’s innocence in the same way as the confession created a belief in his guilt. One explanation for the failure of the Alibi manipulation is that the alibi manipulation was too weak. According to Olson and Wells’ (2004) taxonomy of alibi strength, the cashier’s alibi was strong insofar as she was not familiar with Hines and had no reason to lie on his behalf, but was weak insofar as her statement was not supported by any physical evidence. In the only other study to use alibi evidence to create an expectation of innocence, Dror and Charlton (2006) used a much stronger alibi manipulation -- namely that the suspect was already in police custody when the crime occurred, and thus could not be the culprit. Participants in Study 2 could have discounted the cashier’s alibi on the belief that her recollection of Hines was simply mistaken.

A second, non-mutually exclusive explanation is that the Denial vignette was not neutral, but instead created a belief in Hines’ innocence. Research has found that people exhibit a “truth bias” such that they are naturally inclined to believe the statements of others (Bond & DePaulo, 2006; Levine, Park, & McCormack, 1999). Therefore, our participants may have seen Hines’ denial statement as indicative of innocence. In fact, the tendency to trust self-reports can be so
strong that people sometimes trust self-reports over other objective and independent evidence. For example, Appleby and Kassin (2011) found that jurors sometimes believed confession evidence, a form of self-report, even when it was contradicted by exculpatory DNA.

Given the unique power of confession evidence relative to other forms of evidence (Kassin & Neumann, 1997) and the many variables that impact the persuasiveness of alibi evidence (Olson & Wells, 2004), perhaps a more effective means of inducing an expectation of innocence would have been to tell participants that someone other than the suspect confessed to the crime. This manipulation has been successfully used by others to induce a belief in the suspect’s innocence (see Elaad et al., 1994). Such a manipulation would also help to ensure that innocent-expectancy participants hold a belief in the suspect’s innocence that is equal in strength to guilty-expectancy participants’ belief in the suspect’s guilt.
CHAPTER 11: STUDY THREE METHOD

In Study 2, the use of an evidence lineup increased choosing rates and had a detrimental effect on judgment accuracy. This pattern of results is consistent with research on the relative effects of showups and simultaneous lineups on eyewitness identifications (e.g., Gonzalez et al., 1993). Whereas Study 2 utilized only *simultaneous* evidence lineups, best practice guidelines for eyewitness lineups now recommend presenting lineups *sequentially* when possible (e.g., Wells et al., 1998). Compared to simultaneous lineups, sequential lineups tend to produce a small decline in correct identifications coupled with a substantial decrease in false identifications, and thus tend to produce identifications that are more diagnostic of guilt (Steblay et al., 2001; 2011).

To further explore the analogue between eyewitness and forensic identification, a third study was conducted to test whether the relative effects of sequential versus simultaneous lineup presentations are the same for handwriting evidence as they are for eyewitness identifications. As in Studies 1 and 2, participants served as mock handwriting examiners and analyzed handwriting evidence in the context of a bank robbery investigation. Given the null effect of alibi evidence in Study 2, participants in Study 3 were first told that a suspect had either denied guilt or confessed but then retracted his confession. Participants then compared the suspect’s handwriting sample against a lineup of comparison samples, presented either in simultaneous or sequential format. As in Study 2, some lineups contained a “target” sample that matched the suspect’s sample, while others did not. Study 3 thus employed a 2 (Context: Confession vs. Denial) X 2 (Lineup: Simultaneous vs. Sequential) X 2 (Target: Absent vs. Present) between-subjects design.

**Participants and Design**

A sample of 230 individuals was obtained using Amazon Mechanical Turk (mTurk; see Mason & Suri, 2012). As in all previous studies, participation was restricted to U.S residents
only. Each participant was randomly assigned to one of eight cells produced by the 2 (Context: Denial vs. Confession) X 2 (Lineup: Simultaneous vs. Sequential) X 2 (Target: Present vs. Absent) completely between-subjects design.

Twenty-one participants (9.13%) were excluded after it was determined that they had participated in one of the previous studies. Two participants (0.96%) who indicated that they completed Study 3 on a cellular phone were also excluded. Finally, nine participants (4.35%) were excluded after responding incorrectly to a manipulation check item which asked whether or not the suspect had confessed to the crime. Study 3 was thus based on a sample of $N = 198$, with individual cell sizes ranging from $n = 23$ to $n = 29$ (see Table 6).

The final sample had a mean age of 38.47 ($SD = 12.39$; Range = 18 – 71) and a slight majority of females (58.08%). Two participants (1.01%) reported that they are not currently U.S. citizens. However, all were current U.S. residents, as this was a prerequisite for participation, and the sample included at least one resident from 43 of the 50 U.S. states. With respect to race, most self-identified as White (77.78%), with relatively fewer identifying as Black (9.09%), Asian (5.05%), Hispanic (4.04%), multi-racial (2.02%), and Native American (1.01%). In terms of educational attainment, 30.30% did not hold a college degree, 17.68% held a two-year college degree, 40.91% held a four-year college degree, and 11.11% held a graduate degree.

**Procedure**

The procedure of Study 3 was identical to that of Study 2, with three notable exceptions. First, participants were randomly assigned to read a case summary in which the suspect either confessed or denied guilt (see Appendix E). None of the participants in Study 3 received the case summary in which a restaurant cashier provided an alibi for the suspect.
Second, participants were randomly assigned to view an evidence lineup presented in either a simultaneous or sequential fashion. Those who viewed a simultaneous lineup received the same pre-lineup instructions that were used in the lineup conditions of Study 2 (see Appendix I). They then viewed one of the six simultaneous lineups used in Study 2 (three each in which the target was absent and present; see Appendix J), indicated which, if any, of the robbery notes they believed was written by Mr. Hines, and rated their confidence on a 10-point scale.

Participants who viewed a sequential lineup received a slightly modified form of the pre-lineup instructions used in Study 2, which specified that they would view “several” robbery notes “one-at-a-time” in order to decide which, if any, was written by Mr. Hines. In accordance with best practice guidelines for sequential eyewitness lineups, participants were not told in advance how many robbery notes they would be shown (Lindsay & Wells, 1985; see Appendix I). They then viewed three of the four showups used in Study 2, presented in one of six orders (three each in which the target was absent and present) which corresponded to the left-to-right orders of the six simultaneous lineups (e.g., whereas participants in the simultaneous condition viewed target-absent lineup ACB, those in the sequential condition viewed target-absent showup A, followed by showup C, and then showup B; see Appendix J). For each showup, participants indicated whether they believed that the given robbery note was written by Mr. Hines, and rated their confidence on a 10-point scale. Thus, those who viewed the lineup sequentially made a total of three match judgments, each with its own associated confidence rating.

Third, unlike in Study 2, participants in Study 3 were not asked to explain the rationale for their handwriting judgment(s). After providing match judgments and confidence ratings, participants self-reported the extent to which they felt that their judgments were influenced by the handwriting samples and by the facts of the case, completed a six-item comprehension test to
ensure that they correctly recalled the content of the case summary (the item which asked about Hines’ alibi was not included; see Appendix H), and were fully debriefed.

**Hypotheses**

**H1:** Informed by meta-analytic comparisons of simultaneous and sequential eyewitness lineups by Steblay et al. (2001; 2011), I predicted that sequential evidence lineups would yield lower choosing rates relative to simultaneous evidence lineups.

**H2:** As observed in Study 2, I predicted that participants who were told of the suspect’s confession would demonstrate higher choosing rates than those uninformed of the confession.

**H3:** In light of meta-analyses by Steblay et al. (2001; 2011), I predicted that the effect of Lineup on accuracy would be moderated by Target. When the Target was present, I predicted that sequential lineups would produce fewer correct identifications than simultaneous lineups, but when the Target was absent, sequential lineups would produce fewer false identifications than simultaneous lineups, and that this latter effect would be larger than the former.

**H4:** Again in light of meta-analyses by Steblay et al. (2001; 2011), I predicted that sequential evidence lineups would produce judgments that were more diagnostic of guilt (see Wells & Lindsay, 1980) than those made from simultaneous evidence lineups.

**H5:** Although no such correlation was found in Study 2, given the existence of a weak relationship between eyewitness confidence and accuracy (Sporer et al., 1995), I again predicted a correlation between the accuracy and confidence of handwriting judgments.
CHAPTER 12: STUDY THREE RESULTS

The frequencies of all handwriting judgments across all conditions can be found in Table 6. For target-absent sequential lineups, participants were said to have given a judgment of “cannot be determined” (CBD) only if they judged all three filler samples as CBD. Participants were classified as having made no identification (ID) if they judged at least one of the filler samples as a non-match and did not misidentify any fillers as a match. This group included 21 participants who judged all three fillers as non-matches (70.00% of target-absent non-IDs), five who judged two fillers as non-matches and the third as CBD (16.67%), and four who judged one filler as a non-match and the other two as CBD (13.33%). Participants were classified as having made a filler ID if they misidentified any one or more of the filler samples as a match. This group included three participants who misidentified two filler samples as matches (14.29% of target-absent filler IDs) and one who misidentified all three fillers as matches (4.76%).

For target-present sequential lineups, participants were credited as having made a correct ID if they identified the target sample as a match and did not misidentify either of the two filler samples as a match. Participants were classified as having made a filler ID if they misidentified one or more of the fillers as a match. This group included three participants who misidentified both fillers as matches (15.79% of target-present filler IDs) and seven who identified both the target and one or more fillers as matches (36.84% of target-present filler IDs). Participants were classified as having made a CBD judgment only if they judged all three samples as CBD; this did not occur in this sample. Finally, participants were classified as having made no ID if they identified neither the target nor either of the two fillers as a match.

As in Study 2, I first looked for any systematic trends in CBD judgments as a function of our manipulations. However, CBD judgments were scarce, occurring in only seven participants
(3.54%). They also appeared to be uniformly distributed, as they were found in five of the eight cells, and occurred no more than twice in any cell. As a result, I excluded participants who made CBD judgments from all subsequent analyses (unless otherwise noted).

**Choosing Rates**

Table 7 displays choosing rates (i.e., the proportion of participants who identified any one of the handwriting samples as a match; CBD judgments were considered non-choices) for both Lineup types, broken down by Context and Target. Across all conditions, 56.06% of participants identified one (or more) of the handwriting samples as a match.

Supporting Hypothesis 1, choosing was more frequent overall for simultaneous (65.00%) than for sequential lineups (46.94%), \( \chi^2(1) = 6.56, p = .010, \phi = .18, OR = 2.10 \ [95\% CI: 1.19 – 3.72] \). Follow-up chi-square analyses indicated that this effect was moderated by Target. When the target was absent, choosing occurred more often in simultaneous (68.63%) than sequential (40.38%) lineups, \( \chi^2(1) = 8.28, p = .004, \phi = .28, OR = 3.23 \ [95\% CI: 1.44 – 7.25] \). When the target was present, choosing did not differ between simultaneous (61.22%) and sequential (54.35%) lineups, \( \chi^2(1) = 0.46, p = .497, \phi = .07, OR = 1.33 \ [95\% CI: 0.59 – 3.00] \). The effect of Lineup on choosing was also moderated by Context, such that simultaneous lineups produced more choosing than sequential lineups in the Denial condition, \( \chi^2(1) = 4.67, p = .031, \phi = .21, OR = 2.37 \ [95\% CI: 1.08 – 5.21] \), but Lineup had no effect on choosing in the Confession condition, \( \chi^2(1) = 2.05, p = .152, \phi = .15, OR = 1.83 \ [95\% CI: 0.80 – 4.22] \).

Contrary to Hypothesis 2, Context had no overall effect on choosing, \( \chi^2(1) = 0.90, p = .344, \phi = .07, OR = 1.31 \ [95\% CI: 0.75 – 2.31] \), nor did it affect choosing within simultaneous, \( \chi^2(1) = 0.11, p = .737, \phi = .03, OR = 1.15 \ [95\% CI: 0.51 – 2.63] \), or sequential, \( \chi^2(1) = 0.95, p = .329, \phi = .10, OR = 1.49 \ [95\% CI: 0.67 – 3.31] \), lineups. Context likewise did not affect choosing.
from target-absent, $\chi^2(1) = 0.03, p = .859, \phi = .02, OR = 1.07 [95\% CI: 0.49 – 2.34]$, or target-present lineups, $\chi^2(1) = 1.34, p = .246, \phi = .12, OR = 1.63 [95\% CI: 0.71 – 3.69]$. Within-row comparisons of the frequencies in Table 7 found that Context did not affect choosing in target-absent simultaneous, $\chi^2(1) = 0.10, p = .749, \phi = .05, OR = 1.21 [95\% CI: 0.37 – 3.99]$, target-present simultaneous, $\chi^2(1) = 0.03, p = .858, \phi = .03, OR = 1.11 [95\% CI: 0.35 – 3.51]$, target-absent sequential, $\chi^2(1) = 0.03, p = .870, \phi = .02, OR = 1.10 [95\% CI: 0.36 – 3.36]$, or target-present sequential lineups, $\chi^2(1) = 2.19, p = .139, \phi = .22, OR = 2.44 [95\% CI: 0.74 – 8.01]$. 

I checked for order effects in the three discrete judgments that comprised simultaneous lineups, to ensure that participants were not more likely to identify a given sample as a match solely as a function of its presentation order. Participants who used sequential lineups were categorized in terms of whether they made a positive identification during the first, second, or third round of judgments; non-choosers and those who judged multiple samples as matches were not included. Although there was a slight increase in choosing from the first (24.24%), to second (30.30%), to third (45.45%) round of judgments, a one-way chi-square test confirmed that this difference was not statistically significant, $\chi^2(2) = 2.36, p = .307, \phi = .27$.

**Judgment Accuracy**

To analyze the effects of my manipulations on judgment accuracy, I again dichotomized each participant’s handwriting judgment as either accurate (i.e., correctly identifying the target when it was present, or correctly rejecting a target-absent lineup) or inaccurate (i.e., rejecting a target-present lineup, or misidentifying a filler sample as the target). Observed accuracy rates for each of the eight cells can be found in Table 8. Across all conditions, 31.94% of participants made accurate judgments (with CBD judgments excluded).
Overall, sequential and simultaneous lineups produced accurate judgments 37.11% and 26.60% of the time, respectively. However, this difference did not reach statistical significance, $\chi^2(1) = 2.43, p = .119, \phi = .11, OR = 1.63$ [95% CI: 0.88 – 3.02]. Follow-up chi-square analyses revealed that the effect of Lineup on accuracy was moderated by Target in the manner predicted by Hypothesis 3 (see Figure 7). When the target was present, judgment accuracy did not differ between sequential (13.04%) and simultaneous (26.09%) lineups, $\chi^2(1) = 2.49, p = .115, \phi = .16, OR = 2.35$ [95% CI: 0.80 – 6.94]. When the target was absent, sequential lineups produced more accurate judgments (58.82%) than did simultaneous lineups (27.08%), $\chi^2(1) = 10.14, p = .001, \phi = .32, OR = 3.85$ [95% CI: 1.65 – 8.97]. Context did not moderate the effect of Lineup on accuracy, as the use of a sequential lineup did not improve judgment accuracy in the Confession condition, $\chi^2(1) = 2.26, p = .133, \phi = .16, OR = 1.99$ [95% CI: 0.81 – 4.90] or in the Denial condition, $\chi^2(1) = 0.52, p = .471, \phi = .07, OR = 1.36$ [95% CI: 0.59 – 3.17].

Finally, Context had no overall effect on accuracy, $\chi^2(1) = 0.00, p = .984, \phi = .00, OR = 1.01$ [95% CI: 0.55 – 1.85], and no evidence for moderation by Lineup or Target was found. More specifically, Context did not affect judgment accuracy for sequential lineups, $\chi^2(1) = 0.15, p = .696, \phi = .04, OR = 1.18$ [95% CI: 0.52 – 2.69], simultaneous lineups, $\chi^2(1) = 0.21, p = .651, \phi = .05, OR = 1.24$ [95% CI: 0.49 – 3.11], target-absent lineups, $\chi^2(1) = 0.05, p = .824, \phi = .02, OR = 1.10$ [95% CI: 0.49 – 2.43], or target-present lineups, $\chi^2(1) = 0.28, p = .599, \phi = .06, OR = 1.32$ [95% CI: 0.47 – 3.72]. Similarly, within-row comparisons of the cell frequencies in Table 8 indicated that Context did not affect judgment accuracy for target-absent simultaneous, $\chi^2(1) = 0.39, p = .532, \phi = .09, OR = 1.51$ [95% CI: 0.41 – 5.52], target-present simultaneous, $\chi^2(1) = 0.00, p = 1.00, \phi = .00, OR = 1.00$ [95% CI: 0.27 – 3.73], target-absent sequential, $\chi^2(1) = 0.07, p$
=.788, \varphi = .04, OR = 1.17 [95% CI: 0.38 – 3.59], or target-present sequential, \chi^2(1) = 0.77, p = .381, \varphi = .13, OR = 2.21 [95% CI: 0.36 – 13.47], lineups.

**Accuracy-Confidence Composite**

As in Studies 1 and 2, I created accuracy-confidence scores for participants who viewed simultaneous lineups by combining their dichotomized judgment accuracy with its associated confidence rating, yielding scores that could range from +10 (highly confident accurate judgment) to -10 (highly confident inaccurate judgment). For those who viewed sequential lineups, I first verified that their confidence ratings did not change over time due to order effects. A repeated-measures ANOVA confirmed that confidence ratings did not change between participants’ first (M = 6.47, SD = 2.90), second (M = 6.29, SD = 2.87), and third (M = 6.38, SD = 2.73) round of judgments, F(2,192) = 0.16, p = .857, \eta^2_p = .00. I then calculated the mean of the three confidence ratings that these participants provided (with CBD judgments having been assigned a confidence rating of zero), and similarly combined these mean confidence ratings with their dichotomized accuracy to produce a composite score.

Across all conditions, the sample produced a mean accuracy-confidence score of -2.21 (SD = 6.26). A 2 (Context: Confession vs. Denial) X 2 (Lineup: Simultaneous vs. Sequential) X 2 (Target: Absent vs. Present) on these composite scores revealed no effect of Lineup, F(1,183) = 1.88, p = .172, d = 0.22 [95% CI: -0.67 – 1.10], such that scores were equal for sequential (M = -1.56, SD = 6.57) and simultaneous (M = -2.89, SD = 5.89) lineups. An effect of Target emerged, F(1,183) = 12.07, p = .001, d = 0.52 [95% CI: -0.34 – 1.37], such that scores were higher when the target was absent (M = -0.71, SD = 6.54) than when it was present (M = -3.83, SD = 5.55).

The effects of Target was qualified by a significant Lineup X Target interaction, F(1,183) = 9.51, p = .002, \eta^2_p = .05 (see Figure 8). Providing additional support for Hypothesis 3, simple
effects tests indicated that when the target was absent, accuracy-confidence scores were higher for sequential \((M = 1.16, SD = 6.52)\) than simultaneous \((M = -2.71, SD = 6.01)\) lineups, \(t(97) = 3.07, p = .003, d = 0.62 [95\% CI: -0.60 – 1.85]\). When the target was present, composite scores did not differ between sequential \((M = -4.57, SD = 5.20)\) and simultaneous \((M = -3.09, SD = 5.83)\) lineups, \(t(90) = 1.29, p = .201, d = 0.27 [95\% CI: -0.85 – 1.39]\).

Context had no effect on composite scores, \(F(1,183) = 0.05, p = .830, d = 0.06 [95\% CI: -0.83 – 0.94]\), such that scores did not differ between the Confession \((M = -2.41, SD = 6.13)\) and Denial \((M = -2.04, SD = 6.41)\) conditions. Finally, Context did not interact with either Lineup, \(F(1,183) = 0.06, p = .813, \eta^2_p = .00\), or Target, \(F(1,183) = 0.76, p = .383, \eta^2_p = .00\), and the three-way interaction was not significant, \(F(1,183) = 0.05, p = .825, \eta^2_p = .00\). (Notably, this pattern of results remained identical when those who gave CBD judgments were included.)

**Diagnosticity**

As in Study 2, I calculated *diagnosticity ratios* (Wells & Lindsay, 1980) for simultaneous and sequential lineups both overall and within each Context condition. These can be found in Table 9 along with “percentage guilty” values and statistical tests of the observed diagnosticity ratios against a value of 1.0 (i.e., the point at which correct and false IDs occur with equal frequency; Wells & Olson, 2002). Overall, about a quarter of identifications made from simultaneous (25.53%) and sequential (22.22%) lineups were correct, indicating that both lineup types yielded more false identifications than correct identifications. As such, both lineup types produced poor diagnosticity ratios (0.36 and 0.32, respectively) that were significantly lower than 1.0, \(ps < .005\). Using the arcsine transformation method prescribed by Wells and Olson (2002), I compared the diagnosticity ratios for simultaneous and sequential lineups and found that they did not differ, \(Z = 0.98, p = .327\), thereby failing to support Hypothesis 4.
Judgment Confidence

Contrary to Hypothesis 5, judgment confidence did not predict accuracy, \( r(189) = .04, p = .616 \). Moreover, confidence did not predict accuracy for sequential lineups, \( r(95) = .04, p = .672 \), for simultaneous lineups, \( r(92) = .02, p = .845 \), when a confession was present, \( r(89) = -.08, p = .456 \), when a denial was present, \( r(98) = .15, p = .139 \), when the target was absent, \( r(97) = .05, p = .628 \), or when the target was present, \( r(90) = .06, p = .577 \).

A 2 (Context) X 2 (Lineup) X 2 (Target) ANOVA was performed on participants’ raw confidence scores to explore any overall effects of our manipulations on confidence. Neither an effect of Context, \( F(1,183) = 0.44, p = .506, d = 0.10 \) [95% CI: -0.19 – 0.38], nor of Lineup, \( F(1,183) = 0.12, p = .725, d = 0.06 \) [95% CI: -0.22 – 0.35], nor of Target, \( F(1,183) = 0.70, p = .405, d = 0.12 \) [95% CI: -0.17 – 0.40], was found. None of the two-way interactions – including Context X Lineup, \( F(1,183) = 0.51, p = .478, \eta^2_p = .00 \), Context X Target, \( F(1,183) = 0.09, p = .769, \eta^2_p = .00 \), and Lineup X Target, \( F(1,183) = 0.75, p = .389, \eta^2_p = .00 \), was significant. A marginal three-way interaction emerged, \( F(1,183) = 3.34, p = .069, \eta^2_p = .02 \). However, this interaction was no longer marginal when CBD judgments were included, \( F(1,190) = 1.01, p = .315, \eta^2_p = .01 \), and thus I did not investigate it further.

Self-Reported Influence

As in Studies 1 and 2, a 2 (Context) X 2 (Lineup) X 2 (Target) X 2 (Source: Handwriting vs. Case Facts) mixed ANOVA, with repeated measures on the fourth factor, explored the impact of the manipulations on the extent to which participants felt that their judgments (including CBD judgments) were influenced by the handwriting samples and by the facts of the case. Again, a main effect of Source was found, \( F(1,189) = 96.11, p < .001, d = 0.71 \) [95% CI: 0.55 – 0.87], such that participants reported that their judgments were influenced more by the handwriting (\( M \))
= 7.75, SD = 2.49) than by the facts of the case (M = 4.71, SD = 2.77). Source did not interact with Context, \(F(1,189) = 0.17, p = .685, \eta_p^2 = .00\), Lineup, \(F(1,189) = 0.26, p = .610, \eta_p^2 = .00\), or Target, \(F(1,189) = 0.02, p = .876, \eta_p^2 = .00\), and none of the higher-order interactions was significant, all \(Fs < 2.06\), all \(ps > .150\), all \(\eta_p^2 < .02\).
CHAPTER 13: STUDY THREE DISCUSSION

In Study 3, the manner in which evidence lineups were presented had differential effects on choosing and accuracy as a function of whether or not a matching sample was present. When no matching sample was present, sequential evidence lineups showed a lower choosing rate, and consequently lowered the frequency of false identifications, relative to simultaneous evidence lineups. When a matching sample was present, sequential and simultaneous evidence lineups produced equivalent rates of choosing and of correct identifications. Sequential evidence lineups thus showed a distinct advantage over simultaneous arrays, insofar as they did not impede the ability to identify guilty suspects, but reduced the risk of misidentifying an innocent suspect as guilty – which could ultimately contribute to wrongful conviction. As such, Study 3 showed a “sequential superiority effect” among evidence lineups that closely paralleled the relative effects of sequential and simultaneous eyewitness lineups (Steblay et al., 2001; 2011). As such, Study 3 provided support for the notion that forensic and eyewitness identification are similar processes (e.g., Kassin et al., 2013; Risinger et al., 2002), such that the former can benefit from the vast literature on factors impacting the accuracy of the latter (e.g., Wells et al., 2006).

Unexpectedly, participants who were told of a suspect’s confession did not show higher choosing rates or lower accuracy compared to those who were unaware of the confession. This lack of context effects may be a testament to the value of evidence lineups: When participants viewed an evidence lineup – regardless of how it was presented – the biasing effect of confession evidence was attenuated, such that knowledge of the confession did not diminish accuracy. This is somewhat inconsistent with Study 2, where the confession did not increase choosing but did decrease correct rejections of target-absent simultaneous lineups. As noted above, participants in Study 2 may have been more willing to select the “cannot be determined” option when faced
with a confession and a target-absent lineup, because rejecting the lineup would contradict the confession’s implication that one of the samples matches a guilty suspect. Thus, the effect of confession evidence on judgments made from evidence lineups remains somewhat unclear.
CHAPTER 14: GENERAL DISCUSSION

The current project had two primary aims. Studies 1 and 2 explored whether knowledge of other evidence to suggest a suspect’s guilt or innocence creates systematic bias in judgments of handwriting samples from a suspect and perpetrator. Studies 2 and 3 represented the first empirical tests of how the use of evidence lineups affects accuracy and susceptibility to bias in judgments of handwriting evidence. I will discuss the implications of each of these issues on its own, before turning to the limitations of the current project and avenues for future study.

Exposure to Case Information

In Study 1, participants who were told of the suspect’s confession rated handwriting samples from the suspect and perpetrator as more similar than did those who were unaware of the confession. In Studies 1 and 2, participants who were aware of the suspect’s confession more often misjudged two non-matching handwriting samples as a match. Together, Studies 1 and 2 thus replicated the findings of Kukucka and Kassin (in press) with respect to the impact of confession evidence on handwriting judgments, and thereby added to a growing literature demonstrating that confessions can guide judgments of other evidence in a self-verifying (and erroneous) manner (Dror & Charlton, 2006; Elaad et al., 1994).

Confession evidence is uniquely persuasive (Kassin & Neumann, 1997), but often flawed (Kassin et al., 2010), as countless individuals have been wrongfully convicted after confessing to crimes that they did not commit (e.g., Drizin & Leo, 2004). The practice of providing examiners access to case information has been described as “virtually universal” in some forensic domains (Risinger et al., 2002, p. 32), and thus examiners who are exposed to flawed confession evidence may be prone to giving judgments that corroborate the confession, but -- like the confession -- are flawed. In so doing, examiners provide incriminating evidence that seems to independently
verify the suspect’s guilt, but is in actuality a product of the confession. As a result, the evidence in favor of the suspect’s guilt appears stronger than it truly is (Kassin, 2012).

The results of Study 1 raise important questions as to the exact nature of the process whereby confession evidence impacted handwriting judgments. On the one hand, the fact that confession evidence affected similarity ratings of the two handwriting samples suggests that participants were not merely conforming their match judgments to the confession’s implication of guilt. Basic research has shown that top-down processes can have a fundamental effect on visual perception (e.g., Balcetis & Dunning, 2006; 2010), and thus it is possible that the confession altered the manner in which participants evaluated and perceived the handwriting stimuli. If this is the case, then the finding underscores concerns raised by the National Academy of Sciences (2009) over the ability of confirmation bias to impact forensic examiners without their awareness (see also Dror et al., 2013). Unconscious bias presents a problem that is more difficult to counteract than if erroneous forensic judgments could simply be ascribed to “bad apples” who knowingly produce conclusions that substantiate other evidence (Thompson, 2009). The pernicious effects of confirmation bias are thus unlikely to be eradicated “by mere willpower” if they are in fact altering basic perception (Dror et al., 2013, p. 79).

On the other hand, one could argue that similarity ratings and match judgments were not independent, but rather were calibrated with each other. Participants who gave a high similarity rating may have felt compelled to judge the samples as a match, and vice versa. If this is true, it becomes impossible to discern from the current findings whether the suspect’s confession influenced their perception of the handwriting evidence (i.e., the fundamental manner in which they carried out the handwriting comparison task), their interpretation of it (i.e., their dichotomous judgment of the samples as inculpatory or exculpatory), or both.
Although these data cannot address this question, my paradigm can be adapted in ways that provide some insight into this issue. First, it would be useful to record the amount of time that participants spend engaging in the handwriting comparison task, a measure that would offer some indication of the thoroughness of their analysis. For example, some research suggests that people require less evidence to support an expected conclusion than an unexpected one, and consequently spend less time scrutinizing expected outcomes than unexpected ones (e.g., Ditto, Munro, Apanovitch, Scepansky, & Lockhart, 2003). Therefore, people who expect a certain outcome from the handwriting comparison (i.e., a match or non-match) may quickly find support for this conclusion and spend less time comparing the two samples. One might also make the opposite prediction – namely that people with strong a priori expectations would spend more time comparing the samples, suggesting a greater expenditure of effort to find support for the expected conclusion (e.g., Dror et al., 2012). Either way, systematic differences in reaction time as a function of expectation would be helpful in inferring whether and how biasing expectations affect the fundamental manner in which people analyze the handwriting stimuli.

A second approach would be to track participants’ eye movements during the handwriting comparison task to explore whether differing expectations lead them to focus on different features (e.g., letters) of the same samples. People may selectively seek out features that confirm their pre-existing belief (e.g., Wason, 1960; Darley & Gross, 1983), or they may spend more time focusing on features that refute their expectation (e.g., Ditto et al., 2003). Once again, the results would help to elucidate whether the confession affects global interpretations of the evidence (i.e., match or non-match) in a holistic manner, or whether it affects basic perceptions of the stimuli by causing perceivers to differentially appraise certain features. Such efforts to
understand the process by which contextual information influences judgment will prove indispensable in designing effective means of counteracting the detrimental effects of bias.

In Study 2, I predicted that knowledge of alibi evidence would create an expectation of innocence that would taint handwriting judgments in a comparable (but opposite) manner as the confession had. However, this was not the case. As discussed previously, the alibi used in Study 2 was somewhat weak (and therefore easier to dismiss) insofar as it was not accompanied by physical evidence (Olson & Wells, 2004). In the only other study to test the effect of an alibi on judgments of forensic evidence, Dror and Charlton (2006) used a stronger alibi manipulation, telling fingerprint examiners that the suspect was in police custody when the crime took place. Consequently, one-third of examiners who received this information changed their own prior match judgments such that they now corroborated the suspect’s innocence. It is possible that a similarly strong alibi manipulation would have produced the predicted effects in Study 2.

On the other hand, one could argue that the confession was weak insofar as the suspect recanted and claimed to have been coerced by police, but it nonetheless influenced handwriting judgments in Studies 1 and 2. This pattern is a testament to the power of confession evidence, which is highly persuasive for a number of reasons. First, people tend to under-appreciate the situational pressures of interrogation that lead individuals to falsely confess (Leo & Liu, 2009). Second, people tend to believe that they would never falsely confess, and hence find it difficult to believe that others would (Henkel, Coffman, & Dailey, 2008). Third, people are naturally predisposed to believe the statements of others (Levine et al., 1999), particularly when those statements are against their own self-interest (Levine, Kim, & Blair, 2010). Fourth, confessions -- even when false -- tend to be rich in detail (Appleby et al., 2013; Garrett, 2010).
While confessions create a particularly strong expectation of guilt, Simon’s (2004; 2011) cognitive coherence model posits that any piece of evidence that is considered probative can influence the appraisal of any other piece of evidence (see also Charman, 2013). Consistent with this theory, research has shown that evidentiary judgments can be tainted not only by knowledge of confession evidence, but also by knowledge of eyewitness evidence (Charman et al., 2009; Miller, 1984), alibi evidence (Dror & Charlton, 2006), and indeed any other information that implies the likelihood of a particular outcome (e.g., Bieber, 2012; Nakhaeizadeh et al., in press). Thus, a growing body of literature suggests general caution with respect to exposing forensic examiners to any case information that is not essential for their prescribed analysis.

In contrast, some forensic examiners argue that access to case information “may improve the accuracy of their conclusion” (Elaad, 2013, p. 77). In Study 2, knowledge of a confession did increase correct identifications from target-present showups. Though superficially appealing, this apparent benefit was nothing more than a byproduct of the fact that the confession implied that the samples would match, and it just so happened that they did. In target-absent showups, the confession likewise implied that the samples would match -- but they did not, which produced a corresponding increase in false identifications. Simply put, the fact that the confession improved accuracy for target-present showups was the result of coincidence, not of skill.

The proper role of a forensic examiner is to provide a circumscribed judgment of a single type of evidence for which they have unique expertise. When this independence is compromised, so too is the probative value of their judgment (Page et al., 2012). That is to say, when examiners rely on outside information to produce accurate judgments, their accuracy cannot be attributed to expertise because their conclusions are not the sole product of the evidence at hand (Risinger, 2009). Similarly, Dror et al. (2013) explained that when contextual influences lead examiners to
right conclusion for the wrong reasons, they misrepresent the diagnostic power of their analysis as being stronger than it truly is. Skilled forensic examiners should welcome the opportunity to give their judgments in a context-free environment, as this would enable them to definitively establish their expertise (Wells et al., 2013). Thus, the argument that access to case information benefits examiners appears not only misguided, but counterproductive.

Although participants’ handwriting judgments in Studies 1 and 2 were clearly influenced by their knowledge of the suspect’s confession, they reported that the handwriting evidence had a much stronger impact on their judgments than did the facts of the case. Participants therefore maintained an “illusion of objectivity” with respect to their handwriting judgments (Kunda, 1990), believing them to be the product of a careful and impartial analysis of the handwriting evidence itself. Further, only 1% of participants in Study 2 reported that their knowledge of suspect’s alibi or confession factored into their handwriting judgments. This finding underscores the fact that people often fail to notice and/or appreciate the factors that impact their decision-making (Nisbett & Wilson, 1977) and are unable to recognize their own biases (Pronin, 2007).

If examiners are indeed unaware of how case information impacts their judgments, one might wonder if confirmation bias can be averted simply by educating them about its detrimental effects. Though this remains an open empirical question with respect to the forensic sciences, basic research on bias correction suggests that increasing examiners’ awareness of the potential for bias to affect their judgments would not be an effective intervention (e.g., Wilson et al., 1996). Wilson and Brekke (1994) enumerated four conditions that must be met to successfully curtail cognitive bias: an individual must be aware of the bias, motivated to correct it, aware of the direction and magnitude of the bias, and psychologically capable of correcting it.
Many researchers and legal scholars have sought to raise awareness of how exposure to case information can produce confirmation bias and costly errors in the forensic sciences (e.g., Kassin et al., 2013; Risinger et al., 2002). However, it appears that few examiners are motivated to change their standard procedures to minimize bias. Instead, they maintain that their knowledge of case information is not problematic, claiming that their experience and expertise renders them impervious to the effects of confirmation bias (e.g., Budowle et al., 2009; Elaad, 2013).

Even if examiners are sympathetic to the problem of bias and driven to correct it, they must also be aware of the direction and magnitude of its impact so as to properly calibrate their attempts to counteract its effect on their judgments. To do so would require a great deal of self-insight. In a similar vein, Charman and Wells (2008) had eyewitnesses make an identification from a lineup that, for some, featured one of two “influential variables” (i.e., post-identification feedback or cautionary pre-lineup instructions). Afterward, witnesses estimated the impact that this variable (or lack thereof) had on their judgments. When they then compared the estimated and actual impacts of these variables, Charman and Wells found that witnesses systematically over- or under-estimated how these variables impacted their judgments. Similarly, forensic examiners may over- or under-correct for bias due to lack of insight into its exact magnitude.

In the spirit of “Rosenthal’s rule” (i.e., that scientists be kept “as blind as possible for a long as possible;” Rosenthal, 1978, p. 1007), a more expedient solution would be the adoption of *sequential unmasking* protocols in which the flow of information to examiners is filtered through a case manager who denies them access to information that is extraneous to their analysis (Krane et al., 2008; Mnookin et al., 2011). Sequential unmasking minimizes the likelihood of biasing information leading examiners to erroneous judgments, while increasing the probative value of their judgments by ensuring that they are the sole product of the evidence at hand. Though some
have objected on pragmatic (Budowle et al., 2009) and fiscal (Charlton, 2013) grounds, others have already taken steps to limit examiners’ exposure to extraneous case information and have reported numerous benefits -- and no negative consequences -- as a result (e.g., Found & Ganas, 2013). In short, the observed effects of confession evidence in Studies 1 and 2 add to a rapidly growing literature on confirmation biases in forensic science judgments (Kassin et al., 2013) and thus further bolster arguments in favor of masking protocols (see also Risinger, 2009).

**Forensic Versus Eyewitness Identifications**

Even if examiners are isolated from biasing case information, they may be affected by the base-rate assumption of guilt that is inherent to the standard practice of submitting only one suspect’s sample for testing (Whitman & Koppl, 2010). To counteract this assumption, many researchers (e.g., Kassin et al., 2013; Reese, 2012; Wells et al., 2013) have proposed that examiners should instead compare a given forensic sample against an array of comparison samples (i.e., an *evidence lineup*) in order to determine which (if any) of those samples is a match. Risinger et al. (2002) argued that “the fundamental tasks of the eyewitness and of the forensic examiner share notable similarities” and that evidence lineups would diminish systematic error in the same way that eyewitness lineups do (p. 49).

Taken together, Studies 2 and 3 provide the first empirical data to support the functional similarity of forensic evidence lineups and eyewitness lineups. Study 2 found that simultaneous evidence lineups universally increased choosing rates relative to showups, which increased the frequency of false identifications and decreased overall judgment accuracy. For eyewitnesses, Gonzalez et al. (1993) found that simultaneous lineups likewise elicited more choosing than did showups, such that they led to more false identifications but also more correct identifications. Study 3 then found that sequential evidence lineups produced fewer false identifications than
simultaneous evidence lineups, but did not significantly reduce the rate of correct identifications. Among eyewitnesses, Lindsay and Wells (1985) demonstrated a similar “sequential superiority effect” by showing that sequential eyewitness lineups reduced the rate of false identifications of innocent suspects without diminishing correct identifications of guilty suspects.

A closer look at these effects reveals a striking degree of similarity between the current studies of forensic identification and meta-analyses of eyewitness identification research (Steblay et al., 2003; 2011). Study 2 compared judgments of handwriting evidence in evidence showups and simultaneous evidence lineups, while Steblay et al. (2003) meta-analyzed 12 studies that compared eyewitness identifications from showups and lineups. Table 10 provides the observed differences between showups and simultaneous lineups in direction and magnitude from Study 2 and Steblay et al. (2003). In each, showups enhanced overall identification accuracy relative to simultaneous lineups (by 29% and 21%, in Study 2 and Steblay et al., 2003, respectively), lineups produced more choosing than did showups (by 27%, in each), and simultaneous lineups produced more false identifications than did showups (by 35% and 36%, respectively).

Study 3 compared judgments of handwriting evidence made from sequential versus simultaneous evidence lineups, while Steblay et al. (2001; 2011) have twice meta-analytically compared eyewitness identifications made from simultaneous and sequential lineups. Table 11 compares the direction and magnitude of the observed effects in Study 3 against Steblay et al.’s most recent meta-analysis, and again shows remarkable consistency between the two. For target-present lineups, both found that sequential presentations produced a slight decrease in correct identifications (11% and 14%, respectively), little or no change in filler identifications (4% and 0%, respectively), and a slight increase in the tendency to not make an identification (7% and
14%, respectively). For target-absent lineups, both showed a substantial decrease in false identifications when a sequential presentation is used (29% and 21%, respectively).

The similar effects of evidence and eyewitness lineups on judgment accuracy raises the possibility that other eyewitness phenomena could similarly map onto forensic judgments, such that the adoption of evidence lineups could benefit from the vast existing literature on eyewitness identification practices. For example, Studies 2 and 3 incorporated several admonitions into their pre-lineup instructions that are considered beneficial for eyewitnesses (Malpass & Devine, 1981; Wells, 2006). Also, Studies 2 and 3 randomized the position of the target sample in both types of lineups in light of eyewitness research showing that some positions may be more prone to elicit choosing than others (e.g., Gonzalez et al., 1995; Sporer, 1993).

In other cases, evidence lineups may be immune to some of the problems that plague eyewitness identification. Eyewitness researchers have distinguished between two types of variables that influence eyewitness accuracy: system variables (i.e., aspects of the identification procedure over which the justice system has control) and estimator variables (i.e., variables over which the system has no control, and whose impact is estimated in hindsight; Wells, 1984). Examples of estimator variables include witness stress, cross-race identification, weapon focus, exposure duration, lighting conditions, witness intoxication, criminal disguises, and retention interval (among others; Wells et al., 2006). These uncontrollable factors would be moot for forensic examiners, who would analyze evidence lineups under optimal viewing conditions and without being restricted by the fallibilities of human memory (e.g., Schacter, 2011).

As a practical matter, some researchers have noted that evidence lineups would need include filler samples that are “appropriately similar” to the suspect’s sample (e.g., Whitman & Koppl, 2010, p. 86). Wells et al. (2006) described two possible strategies for selecting filler
photos for eyewitness lineups -- namely, choosing fillers on the basis of their resemblance to the suspect, and choosing fillers that fit the witness’ description of the suspect. They dismissed the former strategy as not viable because it makes the lineup unreasonably difficult, which decreases correct identifications from target-present lineups (Luus & Wells, 1991), and invites a “backfire effect,” whereby the suspect paradoxically stands out in the lineup as the person who was the basis for the filler choices (Wogalter, Marwitz, & Leonard, 1992).

For evidence lineups, selecting fillers on the basis of similarity should not be an issue, given that the lineup will be viewed by experts who are capable of detecting subtle differences between samples, and thus should not be impeded by highly similar fillers in the same way that eyewitnesses are. If the use of highly similar filler samples does produce a high rate of filler identifications, this would lead one to rightly question the competence of the examiner and/or the validity of the technique or discipline (Wells et al., 2013). A more difficult question is how filler samples should be generated. For domains with access to a large database of samples, this should be relatively simple: For example, fingerprint examiners can use AFIS to identify similar prints which can then be used as fillers (Dror et al., 2012). For other domains, lineup construction may be more demanding, but surely not impossible. Risinger et al. (2002) proposed that Evidence and Quality Control officers who are rigorously trained in a given discipline can handle the tasks of obtaining filler samples and screening them for appropriate similarity.

In any event, the results of Studies 2 and 3 suggest an important parallel between the effects of evidence and eyewitness lineups, and thereby lend credence to the claim that evidence lineups can be used to counteract the inherent suggestiveness of testing evidence from only one suspect. Moreover, Wells et al. (2013) explained that the use of evidence lineups would also
allow for the estimation of error rates and inter-examiner reliability, and in so doing, expose fraud and incompetence among examiners, laboratories, or methodologies.

**Limitations and Future Directions**

An important limitation of the current research is that participants were not experts with training and experience in handwriting identification, but rather were untrained participants playing the role of handwriting experts in a mock criminal investigation. Not surprisingly, their lack of expertise was evident in their performance, as they were quite poor at identifying matching handwriting samples when they were present. Indeed, this was the sole exception to the similarity of effects shown in Tables 10 and 11. In Steblay et al. (2003), using a simultaneous lineup increased correct eyewitness identifications by 3%, but in Study 2, simultaneous lineups decreased correct handwriting identifications by 15% (see Table 10).

One would surely expect experts in a given discipline to be more skilled at identifying matching samples than laypeople. However, it does not logically follow that experts are less susceptible to bias by virtue of their training and experience. Indeed, empirical studies have now shown that confirmation bias can impact the judgments of trained examiners across many domains -- including questioned document examination (Miller, 1984), microscopic hair analysis (Miller, 1987), fingerprint examination (Dror & Charlton, 2006), arson investigation (Bieber, 2012), forensic anthropology (Nakhaiezadeh et al., in press), and complex DNA analysis (Dror & Hampikian, 2011) -- and wrongful conviction cases have suggested that contextual factors can produce forensic science errors (e.g., Kassin et al., 2012; Thompson, 2009). Simply put, the position that expertise does not render one immune to confirmation bias appears untenable.

In fact, some have argued that training and experience may actually render experts *more* susceptible to confirmation bias. In a paper on “the paradox of human expertise,” Dror (2011)
described how the acquisition of expertise entails a trade-off between efficiency and flexibility. As individuals gather experience in a given domain, they become more reliant on top-down information in the form of cognitive “shortcuts” (e.g., schemas, information chunking, selective attention) that enable them to process information quickly by focusing their attention on relevant information and ignoring irrelevant information. Similar to decision-making heuristics (Tversky & Kahneman, 1974), experts’ use of top-down shortcuts improves efficiency but can create bias in how bottom-up information is processed, thereby leading to error.

For example, expert fingerprint examiners may largely benefit from selective attention “shortcuts” which focus their attention on important features of fingerprints and allow them to assess prints quickly. However, the process of deciding which features are important may be driven by the expert’s experience and expectations. If an expert expects a suspect to be guilty, he or she may differentially weight the importance of these features as a result of this expectation, such that features that suggest guilt are deemed worthy of attention while those that suggest innocence are overlooked. As experts accumulate more experience, they become more reliant on these top-down shortcuts, while also growing more confident in their abilities as a function of their heightened expertise. Therefore, examiners with the most expertise may paradoxically be the most vulnerable to giving highly confident, erroneous judgments as a result of bias.

Unfortunately, empirical studies of forensic expert performance are quite scarce, and many of those that do exist have been criticized as unscientific, insofar as they were carried out by forensic practitioners who had a vested interest in their outcome (see Risinger & Saks, 2003). Additional high-quality empirical studies of forensic experts are sorely needed, but such studies tend to be logistically difficult to implement, as expert-participants should be unaware that their performance is being measured to ensure that they behave realistically (e.g., Risinger, 2009).
Dror (2009) advised that any attempt to study forensic expert performance be integrated into their normal casework so as not to arouse suspicion and to maximize external validity. However, this requires the cooperation of the supervising agency, which could be difficult to obtain, and researchers would nonetheless need to be sensitive to potential volunteer biases.

Importantly, future studies should attempt to replicate the observed effects among samples of trained and experienced professional examiners in order to test conflicting claims that their expertise renders them more (Dror, 2011) or less (e.g. Butt, 2013) vulnerable to context effects. Research should also be designed to demystify the effects of training and experience by conducting longitudinal studies of professional examiners and/or testing the effects of formal forensic science training on novices. Research on deception detection has shown that experience and certain forms of training increase the confidence with which judgments are made, but fail to improve the accuracy of those judgments (Jordan, Wallace, Kassin, & Hartwig, 2013; Kassin & Fong, 1999; Meissner & Kassin, 2002). Similar empirical assessments of forensic science training protocols would shed light on their efficacy while also testing whether examiners may be prone to overconfidence merely by virtue of having received formal training.

Future research should also attempt to replicate the observed effects across the many diverse domains of forensic science. Although confirmation bias effects have now been demonstrated across a number of different forensic sciences (e.g., Bieber, 2012; Dror & Charlton, 2006; Dror & Hampikian, 2011), others have failed to find such effects in other forensic domains (e.g., Kerstholt et al., 2007; 2010). Accordingly, researchers should be cautious about generalizing results across different forensic science domains. Instead, we must be mindful of the possibility that different forensic science domains may vary in terms of their vulnerability to confirmation bias and the manner in which evidence lineups influence judgments.
Similarly, research should continue to explore the analogue between forensic and eyewitness identification by testing various methods of evidence lineup construction and administration. For example, researchers might explore the effects of changing the number or source of filler samples, of giving examiners feedback on their lineup performance, or of varying the content of pre-lineup instructions. Again, it is possible that the impact of these variables might differ across different forensic science domains.

The National Academy of Sciences (2009) lamented the fact that “research has been sparse on the important topic of cognitive bias in forensic science both regarding their effects and methods for minimizing them” (p. 124) Accordingly, they urged that funding be set aside to “encourage research programs on human observer bias and sources of human error in forensic examinations” (p. 24). While many have now demonstrated the harmful effects of bias on forensic judgments, the current project represents an important first step in studying methods of counteracting the effects of bias. It is my hope that the current findings, though preliminary, will open the door to further study of the costs and benefits of evidence lineup use, so as to fully explore its viability as a means of reducing costly false identifications without diminishing correct ones, and thereby maximizing the probative value of forensic science judgments.
Table 1

*Frequency of Handwriting Judgments by Presentation, Context, and Target (Study 2)*

<table>
<thead>
<tr>
<th></th>
<th>Target Absent</th>
<th></th>
<th>Target Present</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Filler ID No ID CBD</td>
<td>N</td>
<td>Filler ID No ID CBD</td>
</tr>
<tr>
<td>Showup</td>
<td>95</td>
<td>24.21 60.00 15.79</td>
<td>99</td>
<td>35.35 50.51 14.14</td>
</tr>
<tr>
<td>Alibi</td>
<td>33</td>
<td>24.24 69.70 6.06</td>
<td>31</td>
<td>29.03 61.29 9.68</td>
</tr>
<tr>
<td>Denial</td>
<td>29</td>
<td>10.34 72.41 17.24</td>
<td>33</td>
<td>21.21 60.61 18.18</td>
</tr>
<tr>
<td>Confession</td>
<td>33</td>
<td>36.36 39.39 24.24</td>
<td>35</td>
<td>54.29 31.43 14.29</td>
</tr>
<tr>
<td>Lineup</td>
<td>100</td>
<td>59.00 28.00 13.00</td>
<td>96</td>
<td>19.79 35.42 37.50</td>
</tr>
<tr>
<td>Alibi</td>
<td>35</td>
<td>45.71 42.86 11.43</td>
<td>32</td>
<td>18.75 40.62 34.38</td>
</tr>
<tr>
<td>Denial</td>
<td>30</td>
<td>66.67 30.00 3.33</td>
<td>29</td>
<td>17.24 31.03 41.38</td>
</tr>
<tr>
<td>Confession</td>
<td>35</td>
<td>65.71 11.43 22.86</td>
<td>35</td>
<td>22.86 34.29 37.14</td>
</tr>
</tbody>
</table>
Table 2

*Frequency of Choosing by Presentation, Context, and Target (Study 2)*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Overall</th>
<th>Alibi</th>
<th>Denial</th>
<th>Confession</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Showup</strong></td>
<td>194</td>
<td>29.90</td>
<td>26.56&lt;sub&gt;a&lt;/sub&gt;</td>
<td>16.13&lt;sub&gt;a&lt;/sub&gt;</td>
<td>45.59&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Target-Absent</td>
<td>95</td>
<td>24.21</td>
<td>24.24&lt;sub&gt;ab&lt;/sub&gt;</td>
<td>10.34&lt;sub&gt;a&lt;/sub&gt;</td>
<td>36.36&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Target-Present</td>
<td>99</td>
<td>35.35</td>
<td>29.03&lt;sub&gt;a&lt;/sub&gt;</td>
<td>21.21&lt;sub&gt;a&lt;/sub&gt;</td>
<td>54.29&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

| **Lineup**  | 196| 57.14   | 52.24<sub>a</sub> | 57.63<sub>a</sub> | 61.43<sub>a</sub> |
| Target-Absent | 100 | 59.00   | 45.71<sub>a</sub> | 66.67<sub>a</sub> | 65.71<sub>a</sub> |
| Target-Present | 96 | 55.21   | 59.38<sub>a</sub> | 48.28<sub>a</sub> | 57.14<sub>a</sub> |

**Note:** Judgments of “cannot be determined” were considered non-choices. Lineups produced more choosing overall, within both Target conditions, and within the Alibi and Denial conditions (Confession: \( p = .062 \)). Values in the same row not sharing a common subscript differ at \( p < .05 \).
Table 3

*Judgment Accuracy by Presentation, Context, and Target (Study 2)*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Overall</th>
<th>Alibi</th>
<th>Denial</th>
<th>Confession</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Showup</strong></td>
<td>165</td>
<td>55.76</td>
<td>54.24&lt;sub&gt;a&lt;/sub&gt;</td>
<td>54.90&lt;sub&gt;a&lt;/sub&gt;</td>
<td>58.18&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Target-Absent</td>
<td>80</td>
<td>71.25</td>
<td>74.19&lt;sub&gt;ab&lt;/sub&gt;</td>
<td>87.50&lt;sub&gt;b&lt;/sub&gt;</td>
<td>52.00&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Target-Present</td>
<td>85</td>
<td>41.18</td>
<td>32.14&lt;sub&gt;a&lt;/sub&gt;</td>
<td>25.93&lt;sub&gt;a&lt;/sub&gt;</td>
<td>63.33&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td><strong>Lineup</strong></td>
<td>176</td>
<td>26.70</td>
<td>34.43&lt;sub&gt;a&lt;/sub&gt;</td>
<td>25.45&lt;sub&gt;a&lt;/sub&gt;</td>
<td>20.00&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Target-Absent</td>
<td>87</td>
<td>32.18</td>
<td>48.39&lt;sub&gt;a&lt;/sub&gt;</td>
<td>31.03&lt;sub&gt;ab&lt;/sub&gt;</td>
<td>14.81&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Target-Present</td>
<td>89</td>
<td>21.35</td>
<td>20.00&lt;sub&gt;a&lt;/sub&gt;</td>
<td>19.23&lt;sub&gt;a&lt;/sub&gt;</td>
<td>24.24&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

*Note:* Judgments of “cannot be determined” were considered neither accurate nor inaccurate. Showups produced greater accuracy overall, within each Context condition, and within both Target conditions. Values in the same row not sharing a common subscript differ at $p < .02$. 

124
Table 4

*Signal Detection Outcomes with Sensitivity (d’) and Bias (C) Parameters (Study 2)*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Hit</th>
<th>Miss</th>
<th>False Alarm</th>
<th>Correct Rejection</th>
<th>d’</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Showup</td>
<td>165</td>
<td>.41</td>
<td>.59</td>
<td>.29</td>
<td>.71</td>
<td>0.34</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alibi</td>
<td>59</td>
<td>.32</td>
<td>.68</td>
<td>.26</td>
<td>.74</td>
<td>0.19</td>
<td>0.56</td>
</tr>
<tr>
<td>Denial</td>
<td>51</td>
<td>.26</td>
<td>.74</td>
<td>.13</td>
<td>.87</td>
<td>0.50</td>
<td>0.90</td>
</tr>
<tr>
<td>Confession</td>
<td>55</td>
<td>.63</td>
<td>.37</td>
<td>.48</td>
<td>.52</td>
<td>0.39</td>
<td>-0.15</td>
</tr>
<tr>
<td>Lineup</td>
<td>176</td>
<td>.21</td>
<td>.79</td>
<td>.68</td>
<td>.32</td>
<td>-1.26</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alibi</td>
<td>61</td>
<td>.20</td>
<td>.80</td>
<td>.52</td>
<td>.48</td>
<td>-0.88</td>
<td>0.40</td>
</tr>
<tr>
<td>Denial</td>
<td>55</td>
<td>.19</td>
<td>.81</td>
<td>.69</td>
<td>.31</td>
<td>-1.39</td>
<td>0.19</td>
</tr>
<tr>
<td>Confession</td>
<td>60</td>
<td>.24</td>
<td>.76</td>
<td>.85</td>
<td>.15</td>
<td>-1.74</td>
<td>-0.17</td>
</tr>
</tbody>
</table>

*Note:* Judgments of “cannot be determined” were excluded. Larger d’ values indicate greater discrimination ability. Negative values of C indicate bias toward positive responses.
Table 5

Conditional Probabilities of Correct and False IDs, Diagnosticity (dx) Ratios, and Percentages

Guilty by Presentation and Context (Study 2)

|                | p(Correct ID | TP) | p(False ID | TA) | dx Ratio | % Guilty | Z         | p         |
|----------------|-------------|------|-----------|---------|----------|-----------|-----------|
| Showup         | .3535       | .2421| 1.46      | 60.34   | 1.70     | < .001    |           |
| Alibi          | .2903       | .2424| 1.20      | 52.94   | 0.43     |           | .665      |
| Denial         | .2121       | .1034| 2.05      | 70.01   | 1.19     |           | .235      |
| Confession     | .5428       | .3636| 1.49      | 61.29   | 1.49     |           | .136      |

|                | p(Correct ID | TP) | p(False ID | TA) | dx Ratio | % Guilty | Z         | p         |
|----------------|-------------|------|-----------|---------|----------|-----------|-----------|
| Lineup         | .1979       | .5900| 0.34      | 24.36   | -5.81    | < .001    |           |
| Alibi          | .1875       | .4571| 0.41      | 27.27   | -2.41    | .016      |           |
| Denial         | .1724       | .6667| 0.26      | 20.00   | -4.05    | < .001    |           |
| Confession     | .2286       | .6571| 0.35      | 25.81   | -3.74    | < .001    |           |

Note: Judgments of “cannot be determined” were included. Z values reflect standardized differences between observed diagnosticity ratios and 1.0. All p values are two-tailed.
Table 6

*Frequency of Handwriting Judgments by Lineup, Context, and Target (Study 3)*

<table>
<thead>
<tr>
<th>Lineup</th>
<th>Target Absent</th>
<th>Target Present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ID</td>
<td>No ID</td>
</tr>
<tr>
<td>Simultaneous</td>
<td>51</td>
<td>68.63</td>
</tr>
<tr>
<td>Denial</td>
<td>27</td>
<td>66.67</td>
</tr>
<tr>
<td>Confession</td>
<td>24</td>
<td>70.83</td>
</tr>
<tr>
<td>Sequential</td>
<td>52</td>
<td>40.38</td>
</tr>
<tr>
<td>Denial</td>
<td>29</td>
<td>41.38</td>
</tr>
<tr>
<td>Confession</td>
<td>23</td>
<td>39.13</td>
</tr>
</tbody>
</table>
Table 7

*Frequency of Choosing by Lineup, Context, and Target (Study 3)*

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th>Overall</th>
<th>Denial</th>
<th>Confession</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simultaneous</strong></td>
<td>100</td>
<td>65.00</td>
<td>63.46</td>
<td>66.67</td>
</tr>
<tr>
<td>Target-Absent</td>
<td>51</td>
<td>68.63</td>
<td>66.67</td>
<td>70.83</td>
</tr>
<tr>
<td>Target-Present</td>
<td>49</td>
<td>61.22</td>
<td>60.00</td>
<td>62.50</td>
</tr>
<tr>
<td><strong>Sequential</strong></td>
<td>98</td>
<td>46.94</td>
<td>42.31</td>
<td>52.17</td>
</tr>
<tr>
<td>Target-Absent</td>
<td>52</td>
<td>40.38</td>
<td>41.38</td>
<td>39.13</td>
</tr>
<tr>
<td>Target-Present</td>
<td>46</td>
<td>54.35</td>
<td>43.48</td>
<td>65.22</td>
</tr>
</tbody>
</table>

**Note:** Judgments of “cannot be determined” were considered non-choices.
Table 8  

*Judgment Accuracy by Lineup, Context, and Target (Study 3)*

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th>Overall</th>
<th>Denial</th>
<th>Confession</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simultaneous</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target-Absent</td>
<td>48</td>
<td>27.08</td>
<td>30.77</td>
<td>22.73</td>
</tr>
<tr>
<td>Target-Present</td>
<td>46</td>
<td>26.09</td>
<td>26.09</td>
<td>26.09</td>
</tr>
<tr>
<td><strong>Sequential</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target-Absent</td>
<td>51</td>
<td>58.82</td>
<td>57.14</td>
<td>60.87</td>
</tr>
<tr>
<td>Target-Present</td>
<td>46</td>
<td>13.04</td>
<td>8.70</td>
<td>17.39</td>
</tr>
</tbody>
</table>

*Note:* Judgments of “cannot be determined” were considered neither accurate nor inaccurate.
Table 9

*Conditional Probabilities of Correct and False IDs, Diagnosticity (dx) Ratios, and Percentages

Guilty by Lineup and Context (Study 3)

<table>
<thead>
<tr>
<th></th>
<th>p(Correct ID</th>
<th>TP)</th>
<th>p(False ID</th>
<th>TA)</th>
<th>dx Ratio</th>
<th>% Guilty</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simultaneous</strong></td>
<td>.2449</td>
<td>.6863</td>
<td>0.36</td>
<td>25.53</td>
<td>-4.59</td>
<td>&lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denial</td>
<td>.2400</td>
<td>.6667</td>
<td>0.36</td>
<td>25.00</td>
<td>-3.19</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confession</td>
<td>.2500</td>
<td>.7083</td>
<td>0.35</td>
<td>26.09</td>
<td>-3.30</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sequential</strong></td>
<td>.1304</td>
<td>.4038</td>
<td>0.32</td>
<td>22.22</td>
<td>-3.15</td>
<td>.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denial</td>
<td>.0870</td>
<td>.4138</td>
<td>0.21</td>
<td>14.29</td>
<td>-2.86</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confession</td>
<td>.1739</td>
<td>.3913</td>
<td>0.44</td>
<td>30.77</td>
<td>-1.67</td>
<td>.096</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* Judgments of “cannot be determined” were included. Z values reflect standardized differences between observed diagnosticity ratios and 1.0. All p values are two-tailed.
Table 10

Comparison of Steblay et al. (2003) Meta-Analysis and Study 2 Results

<table>
<thead>
<tr>
<th></th>
<th>Steblay et al. (2003)</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Showup (%)</td>
<td>Sim (%)</td>
</tr>
<tr>
<td>Overall Accuracy</td>
<td>69</td>
<td>48</td>
</tr>
<tr>
<td>Overall Choosing</td>
<td>27</td>
<td>54(^a)</td>
</tr>
<tr>
<td><strong>Target Present</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing</td>
<td>46</td>
<td>71(^a)</td>
</tr>
<tr>
<td>Correct ID</td>
<td>47</td>
<td>50</td>
</tr>
<tr>
<td>Miss</td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td><strong>Target Absent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing</td>
<td>15</td>
<td>43(^a)</td>
</tr>
<tr>
<td>Correct Rejection</td>
<td>85</td>
<td>49</td>
</tr>
<tr>
<td>Filler ID</td>
<td>15</td>
<td>51</td>
</tr>
</tbody>
</table>

**Note:** The “Miss” (Target Present) and “Correct Rejection” (Target Absent) categories include judgments of “I don’t know” (Steblay et al., 2003) and “cannot be determined” (Study 2).

\(^a\) Steblay et al. (2003) do not report choosing rates separately for simultaneous and sequential lineups; these values represent choosing rates collapsed across both types of lineups.
Table 11

Comparison of Steblay et al. (2011) Meta-Analysis and Study 3 Results

<table>
<thead>
<tr>
<th></th>
<th>Steblay et al. (2011)</th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sim (%)</td>
<td>Seq (%)</td>
</tr>
<tr>
<td><strong>Target Present</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct ID</td>
<td>52</td>
<td>38</td>
</tr>
<tr>
<td>Filler ID</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>No Choice</td>
<td>27</td>
<td>41</td>
</tr>
<tr>
<td><strong>Target Absent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Rejection</td>
<td>43</td>
<td>64</td>
</tr>
<tr>
<td>Filler ID</td>
<td>57</td>
<td>36</td>
</tr>
</tbody>
</table>

*Note:* The “No Choice” and “Correct Rejection” categories include judgments of “I don’t know” (Steblay et al., 2011) and “cannot be determined” (Study 3).
Figure 1

* Means differ at $p < .01$
Figure 2

Confession X Source Interaction on Self-Reported Influence (Study 1)
Figure 3

*Context X Target Interaction on Judgment Accuracy (Study 2)*

![Chart showing the interaction between context and target on judgment accuracy. The chart compares response accuracy for target present and target absent conditions across alibi, denial, and confession contexts.](chart_image)
Figure 4

*Context X Target Interaction on Accuracy-Confidence Composite Scores (Study 2)*
Figure 5

*Context X Source Interaction on Self-Reported Influence (Study 2)*
Figure 6

Context X Target Interaction on Number of Handwriting Features Cited in Open-Ended Explanation for Handwriting Judgment (Study 2)
Figure 7

Lineup X Target Interaction on Judgment Accuracy (Study 3)
Figure 8

Lineup X Target Interaction on Accuracy-Confidence Composite Scores (Study 3)
Appendix A

Informed Consent

You are invited to participate in a research study of how individuals assess evidence collected during a criminal investigation. This page contains important information about this study. Please read all of the information on this page before agreeing to participate.

We hope to recruit a total of 160\(^1\) individuals to participate. If you agree to participate, you will assume the role of a forensic examiner who has been hired to assist with a criminal investigation. You will receive information about an actual crime, and then analyze and provide opinions about several pieces of evidence.

This study should take about 15-20 minutes to complete. If at any point you wish to stop participating in this study, you are free to do so. In exchange for completing this study, you will receive $0.50, credited to your Amazon Mechanical Turk account.

**Note:** After completing this study, you will be given a verification code to indicate that you have completed it. Be sure to write this code down! In order to receive payment, you must enter this code into Mechanical Turk. You cannot be paid without entering this code.

Your participation in this study will provide you with an opportunity to learn about how psychology can inform the legal system. Participation in this study is completely voluntary and refusal to participate or withdraw from this study will not result in any adverse consequences.

Any information obtained in this study will be kept strictly confidential. In any written reports or publications, no one participant will be identified or identifiable, and only aggregate data will be presented. Research records will be kept in a password-protected computer file; only the researchers will have access to these records. All data will be destroyed five years after the publication of the study, and no report of the data will ever be linked to you.

This study has been approved by the John Jay College of Criminal Justice Institutional Review Board (Study #468525-2). The researcher conducting this study is Jeff Kukucka, a doctoral candidate at John Jay College. If you have any questions, you may contact him at jkukucka@jjay.cuny.edu.

If you have any concerns or questions about the rights of research participants or the ethics of this study, please contact the John Jay College Human Research Protection Program (HRPP) Office, at jj-irb@jjay.cuny.edu, or (212) 237-8961.

By checking the box below and clicking “next,” you are indicating that you have read and understood the information on this page, and that you agree to participate in this study.

[ ] I have read all of the information on this page and I agree to participate in this research study.\(^2\)

---

\(^1\) In Study 2, this number was changed to 480. In Study 3, it was changed to 240.

\(^2\) Participants had to check a box to indicate consent in order to proceed to the next page.
Appendix B

Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984)

Please indicate the extent to which you agree or disagree with each of the following statements using the scale below.

```
<table>
<thead>
<tr>
<th></th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very strongly disagree</td>
<td>Strongly disagree</td>
<td>Moderately disagree</td>
<td>Slightly disagree</td>
<td>Neither agree nor disagree</td>
<td>Slightly agree</td>
<td>Moderately agree</td>
<td>Strongly agree</td>
<td>Very strongly agree</td>
</tr>
</tbody>
</table>
```

1) I would prefer complex problems to simple problems.

2) I like to have the responsibility of handling a situation that requires a lot of thinking.

*3) Thinking is not my idea of fun.

*4) I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

*5) I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.

6) I find satisfaction in deliberating hard and for long hours.

*7) I only think as hard as I have to.

*8) I prefer to think about small, daily projects rather than long-term ones.

*9) I like tasks that require little thought once I’ve learned them.

10) The idea of relying on thought to make my way to the top appeals to me.

11) I really enjoy a task that involves coming up with new solutions to problems.

*12) Learning new ways to think doesn’t excite me very much.

13) I prefer my life to be filled with puzzles that I must solve.

14) The notion of thinking abstractly is appealing to me.
15) I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

*16) I feel relief rather than satisfaction after completing a task that required a lot of mental effort.

*17) It is enough for me that something gets the job done; I don’t care how or why it works.*

18) I usually end up deliberating about issues even when they do not affect me personally.

* = Item is reverse-scored.
Appendix C

Demographic Questionnaire

Please answer the following questions about yourself:

1) What is your gender?
   Male
   Female

2) What is your age? ________

3) Are you currently a U.S. citizen?
   Yes
   No

4) In which U.S. state do you currently reside?
   [All 50 U.S. states listed as response options.]

5) Which of the following best describes your racial or ethnic background?
   White / Caucasian
   African-American
   Hispanic / Latino
   Asian / Pacific Islander
   Native American
   More than one of these
   Other

6) What is the highest level of education that you have completed?
   Less than High School
   High School Diploma / GED
   Two-Year College (Associates Degree)
   Four-Year College (Bachelors Degree)
   Masters Degree (MA, MS, etc.)
   Doctoral / Professional Degree (PhD, JD, MD)
7) What is your marital status? (*Study 1 only*)

- Single (Never Married)
- Married
- Separated
- Widowed
- Divorced

8) In which political party are you registered to vote? (*Study 1 only*)

- Republican Party
- Democratic Party
- Other (e.g., Green, Libertarian, etc.)
- None / Unaffiliated / Not Registered

9) Have you ever traveled outside of the United States? (*Study 1 only*)

- Yes
- No

10) Are you right-handed or left-handed? (*Study 1 only*)

- Right-handed
- Left-handed

7) On what kind of device are you completing this study? (*Studies 2 and 3 only*)

- Desktop Computer
- Laptop Computer
- Tablet
- Cell Phone
- Other
Appendix D
Mock Examiner Instructions

**Study 1:**

For the remainder of this study, you will assume the role of a **handwriting identification expert**. Your job is to analyze and compare handwriting samples, and to give your opinion as to whether or not the samples were written by the same person.

You are hired by police investigators to analyze handwriting evidence because your opinions are often useful in their efforts to solve crimes.

The next page contains a letter from the local police department, along with information about an actual bank robbery case that has been provided by investigators who are working on this case. These investigators have asked that you first review the facts of this case, and then examine handwriting evidence that is relevant to the case, and report your opinions back to them.

If you understand what you have been asked to do, please check the box and continue to the next page, which contains a summary of this case provided by police investigators.

**Studies 1a, 2, and 3:**

For the remainder of this study, you will assume the role of a **handwriting identification expert**. Your job is to analyze and compare handwriting samples, and to give your opinion as to whether or not the samples were written by the same person.

To do this, handwriting experts look for similarities between handwriting samples, while also keeping in mind that no one’s handwriting is always the same from one occasion to another. Even two samples written by the same person will not be perfectly identical.

Imagine that you are hired by police who are investigating a crime. They ask you to look at several handwriting samples and to give your personal opinions as to whether you think they were written by the same person.

The next page contains a letter from the local police department, along with information about an actual bank robbery case that has been provided by investigators who are working on this case. These investigators have asked that you first review the facts of this case, and then examine handwriting evidence that is relevant to the case, and report your opinions back to them.

If you understand what you have been asked to do, please check the box and continue to the next page, which contains a summary of this case provided by police investigators.
Appendix E

Case Summary

ALL Conditions:

On December 10, 2008, a young African American male entered the Broadway National Bank on 457 Broadway, in Chelsea, Massachusetts, at approximately 10:30am.

The man approached one of the tellers, Ms. Jeanne Dunne, a Caucasian female, age 26, and handed her a note explaining that he was armed with a gun and ordering her to produce a large sum of cash. This hand-written note has been retained as evidence. The man then opened his jacket to reveal that he was carrying a silver handgun.

Ms. Dunne handed the robber $10,982 in cash from her drawer. He stuffed the money into his jacket and fled. As he exited, Ms. Dunne sounded the alarm, thereby alerting police, who arrived at 10:50am.

When interviewed by police, Ms. Dunne described the robber as a black male, around 6-feet, 2-inches tall, wearing a heavy black jacket and jeans. When asked to estimate his age, Ms. Dunne said he was “maybe about 35, but it was hard to tell.” A customer who was in the bank at the time but not paying attention agreed but could not be more specific.
A surveillance photo from the bank’s security system (see below) confirmed Ms. Dunne’s general description of a black male wearing a heavy black coat.

Police immediately searched the neighborhood for anyone who fit the description. At 11:00am, they stopped a red Honda Accord, less than a half-mile from the bank, driving over the posted speed limit. The vehicle owner and driver, Johanna Hines, a black male, age 39, wore a dark-colored coat and jeans and fit the age and height description given by Ms. Dunne.

Mr. Hines was stopped and questioned regarding his whereabouts that morning. The interviewing officer reported that Hines behaved suspiciously at this time, avoiding eye contact and looking nervous. The officer asked Hines to exit the vehicle so that a search could be performed. Police found nothing in this search -- neither the gun nor the stolen cash.

One week after the robbery, Ms. Dunne (the bank teller) was shown a photo lineup of six individuals who fit her description, one of which was Mr. Hines. When asked to identify the man who had robbed the bank, she picked Johanna Hines, though she admitted that she was "not 100% sure that was the guy.”

At that point, Hines was picked up by police, read his *Miranda* rights, and brought into the police station for questioning. Prior to being questioned, Mr. Hines waived his *Miranda* rights and agreed to speak with police. A copy of Mr. Hines’ hand-written Miranda waiver has been retained as evidence.

**Confession Condition ONLY:**

After being questioned by detectives for nearly 3 hours, Mr. Hines confessed to robbing the Broadway National Bank. Hines admitted that he purchased a silver 9mm pistol "off the street" to use in the robbery, wrote a note for the teller, and rubbed Vaseline on his face so that it would be difficult to see his face in the security cameras. He explained that he owed money and was out of work. Mr. Hines said he drove to the bank and parked near a dumpster around the corner, entered with the gun hidden in his jacket, and approached a teller who was wearing a white shirt.
and a red scarf. When asked what he did with the money, Hines said that he hid the money and the gun in the dumpster before driving off, and that he had planned to retrieve them the next day, but when he returned they were gone.

A copy of Mr. Hines’ signed statement is shown below:

---

On December 10, 2008, I stole $10,982 from the Broadway National Bank in Chelsea. I bought a 9mm gun on the street and hid it in my jacket. When I went in the bank I gave the teller a note asking for money and I showed her my gun. Then she gave me the money and I ran out. I stole the money because I needed to pay off my debt. I hid the gun and money in a dumpster so I could come back and get it later, but when I came back it was gone.

---

After meeting with a lawyer, Hines recanted his confession, and is now claiming that he was coerced into signing it by the detectives who questioned him.

---

**Denial Condition ONLY:**

After being questioned by detectives for nearly 3 hours, Mr. Hines still denied having any knowledge of or involvement in the robbery. Hines stated that on the morning of the robbery, he drove to the nearby McDonald’s restaurant on 80 Broadway, parked near a dumpster around the corner, ordered a large coffee and a breakfast sandwich from a cashier who was wearing a white shirt and a red scarf, and ate breakfast alone while reading the newspaper. Mr. Hines estimated that he was in the restaurant for approximately 45 minutes. Maintaining his innocence throughout the session, Hines said that he was stopped by police on his way home from McDonald’s that morning. If he seemed nervous, he said, it was because police shouted at him to get out of the car and he did not know why.

A copy of Mr. Hines’ signed statement is shown below:
Alibi Condition ONLY: (Study 2 only)

After being questioned by detectives for nearly 3 hours, Mr. Hines still denied having any knowledge of or involvement in the robbery. Hines stated that on the morning of the robbery, he had breakfast alone at the nearby McDonald’s restaurant on 80 Broadway. Maintaining his innocence throughout the session, Hines said that he was stopped by police on his way home from McDonald’s that morning. If he seemed nervous, he said, it was because police shouted at him to get out of the car and he did not know why.

Police later interviewed the cashier who was working at McDonald’s on the morning of the robbery. When shown a photo of Hines, she said that she remembered seeing him on the morning of the robbery. She stated that he was her first customer during her shift which started at 10am. The cashier explained that Hines ordered a coffee and a breakfast sandwich, read the newspaper alone, and stayed for approximately 45 minutes.

A copy of the cashier’s signed statement is shown below:

On December 10, 2008, I ate breakfast at the McDonalds restaurant in Chelsea. I ordered a coffee and a breakfast sandwich from a lady at the cashier and then I ate breakfast alone. I stayed there and read the newspaper for about 45 minutes and then went home. On the way home I was pulled over by police for speeding. I had nothing to do with the robbery of Broadway National Bank because I was at McDonalds when the robbery happened.
On December 10, 2008, I was working as a cashier at the McDonalds restaurant in Chelsea. I remember seeing Mr. Hines that morning because he was my first customer during my shift which started at 10am. He ordered a coffee and a breakfast sandwich, and then ate alone. He stayed and read the newspaper for about 45 minutes and then he left. He could not have robbed Broadway National Bank because he was at McDonalds when it happened.
Appendix F

Handwriting Samples (Study 1)³

Low-Similarity Pair:

Exhibit A:  
I have a gun  
keep quiet or I will  
shoot you  
Give me all your cash!

Exhibit B:  
I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

High-Similarity Pair: (also used in Study 1a)

Exhibit A:  
I have a gun  
keep quiet or I will  
shoot you  
Give me all your cash!

Exhibit B:  
I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

³ These stimuli were larger when shown to participants; they have been reduced to fit on this page.
Appendix G

Memo from Investigators (Study 1)

Dear Sir or Madam,

Thank you for providing your opinions on these two items of handwriting evidence that we have collected as part of our investigation of the armed robbery of the Broadway National Bank.

With tasks such as these, we’ve found that individuals tend to make better judgments when they are given more time to consider and analyze the evidence. Because we want to be sure that we get this right, we would like you to have another look at the facts of the case and the handwriting samples, and to produce a final set of opinions for us. We look forward to receiving your opinions regarding this evidence.

Sincerely,

[Signature]

Sgt. J. O’Connor
Chelsea Police Department
Appendix H

Comprehension Test

Please answer the following questions to demonstrate that you have understood the details of this case.

1) What is the crime being investigated in this case?
   - Automobile theft
   - Murder
   - Armed robbery
   - Drug possession

2) Who provided a physical description of the culprit?
   - A customer
   - A bank teller
   - A security guard
   - A police officer

3) What did the police find in Mr. Hines’ car?
   - A gun
   - A large sum of cash
   - A ski mask
   - None of these

4) When shown a photo lineup, did the eyewitness identify Mr. Hines as the culprit?
   - Yes
   - No

5) When questioned by police, did Mr. Hines confess to the crime?
   - Yes
   - No

6) Did Mr. Hines appear to know the details of the crime? (Studies 2 and 3 only)
   - Yes
   - No

7) Could anyone vouch for Mr. Hines’ whereabouts during the time when the crime occurred? (Study 2 only)
   - Yes
   - No
Appendix I

Pre-Lineup Instructions (Studies 2 and 3)

Showup Condition:

Dear Sir or Madam,

You have been asked to examine handwriting samples that are relevant to our investigation of the armed robbery of Broadway National Bank. These samples include:

Exhibit A: The hand-written form on which Mr. Hines agreed to waive his rights before speaking to police, which reads, "I understand my rights to remain silent and to call a lawyer and I agree to talk at this time."

Exhibit B: The hand-written note that the robber gave to the bank teller, which reads, "I have a gun. Keep quiet or I will shoot you. Give me all your cash."

Please examine these samples, and provide your opinion as to whether the robbery note was written by Mr. Hines. In doing so, please keep the following things in mind...

• The note used in the robbery may or may not have been written by Mr. Hines.
• It is just as important to clear innocent persons from suspicion as it is to identify the guilty.
• A person’s handwriting may not look exactly the same on two different occasions.

Regardless of whether or not you are able to give a definitive opinion about these samples, we will continue to investigate this incident. We look forward to receiving your opinion.

Sincerely,

[Signature]

Sgt. J. O’Connor
Chelsea Police Department
Lineup Condition (Study 2) / Simultaneous Lineup Condition (Study 3):

Dear Sir or Madam,

You have been asked to examine handwriting samples that are relevant to our investigation of the armed robbery of Broadway National Bank. These samples include:

**Exhibit A:** The hand-written form on which Mr. Hines agreed to waive his rights before speaking to police, which reads, "I understand my rights to remain silent and to call a lawyer and I agree to talk at this time."

**Exhibits B-D:** Three hand-written notes, which read, "I have a gun. Keep quiet or I will shoot you. Give me all your cash." One of these three notes (Exhibit B, C, or D) was used in the robbery of Broadway National Bank.

Please examine these samples, and provide your opinion as to which -- if any -- of the robbery notes was written by Mr. Hines. In doing so, please keep the following things in mind:

- It is possible that none of these three robbery notes was written by Mr. Hines.
- It is just as important to clear innocent persons from suspicion as it is to identify the guilty.
- A person’s handwriting may not look exactly the same on two different occasions.

Regardless of whether or not you are able to give a definitive opinion about these samples, we will continue to investigate this incident. We look forward to receiving your opinion.

Sincerely,

[Signature]

Sgt. J. O’Connor
Chelsea Police Department
Sequential Lineup Condition (Study 3):

Dear Sir or Madam,

You have been asked to examine handwriting samples that are relevant to our investigation of the armed robbery of Broadway National Bank. These samples include:

**Exhibit A:** The hand-written form on which Mr. Hines agreed to waive his rights before speaking to police, which reads, "I understand my rights to remain silent and to call a lawyer and I agree to talk at this time."

**Exhibits B:** Several hand-written notes, which read, "I have a gun. Keep quiet or I will shoot you. Give me all your cash." One of these notes was used in the robbery of Broadway National Bank.

Please examine these robbery notes one-at-a-time, and provide your opinion as to which -- if any -- of the robbery notes was written by Mr. Hines. In doing so, please keep the following things in mind:

- It is possible that none of these robbery notes was written by Mr. Hines.
- It is just as important to clear innocent persons from suspicion as it is to identify the guilty.
- A person's handwriting may not look exactly the same on two different occasions.

Regardless of whether or not you are able to give a definitive opinion about these samples, we will continue to investigate this incident. We look forward to receiving your opinion.

Sincerely,

Sgt. J. O'Connor
Chelsea Police Department
Appendix J

Evidence Showups and Lineups

Target-Absent Showup (Version A):

Exhibit A:

Exhibit B

---

4 These stimuli were larger when shown to participants; they have been reduced to fit on this page.
Target-Absent Showup (Version B):

Exhibit A:

I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

Exhibit B:

I have a gun.
Keep quiet or I will shoot you.
Give me all your cash!
Target-Absent Showup (Version C):

Exhibit A:

I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

Exhibit B

I have a gun.
Keep quiet or I will shoot you.
Give me all your cash!
Target-Present Showup:

Exhibit A:

Chelsea Police
19 Park Street, Chelsea, MA. 02150

I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

Exhibit B

I have a gun.
Keep quiet or I will shoot you.
Give me all your cash!
Target-Absent Lineup (Version ACB):

Exhibit A:

I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

Exhibit B

Exhibit C

Exhibit D
Target-Absent Lineup (Version BAC):

Exhibit A:

I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

Exhibit B

I have a gun. Keep quiet or I will shoot you.

Exhibit C

I have a gun. Keep quiet or I will shoot you.

Exhibit D

I have a gun. Give me all your cash!

Give me all your cash!
Target-Absent Lineup (Version CBA):

Exhibit A:

I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

Exhibit B
I have a gun.
Keep quiet or I will shoot you.
Give me all your cash!

Exhibit C
I have a gun.
Keep quiet or I will shoot you.
Give me all your cash!

Exhibit D
I have a gun.
Keep quiet or I will shoot you.
Give me all your cash!
Target-Present Lineup (Version TCB):

Exhibit A:

I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

<table>
<thead>
<tr>
<th>Exhibit B</th>
<th>Exhibit C</th>
<th>Exhibit D</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have a gun.</td>
<td>I have a gun.</td>
<td>I have a gun.</td>
</tr>
<tr>
<td>Keep quiet or I will shoot you</td>
<td>Keep quiet or I will shoot you</td>
<td>Keep quiet or I will shoot you</td>
</tr>
<tr>
<td>Give me all your cash!</td>
<td>Give me all your cash!</td>
<td>Give me all your cash!</td>
</tr>
</tbody>
</table>
Target-Present Lineup (Version CTA):

Exhibit A:

I understand my rights to remain silent and to call a lawyer and I agree to talk at this time.

Exhibit B: I have a gun. Keep quiet or I will shoot you. Give me all your cash!
Exhibit C: I have a gun. Keep quiet or I will shoot you. Give me all your cash!
Exhibit D: I have a gun. Keep quiet or I will shoot you. Give me all your cash!
Target-Present Lineup (Version ABT):

Exhibit A:

Exhibit B

Exhibit C

Exhibit D


Dupuis, P. R., & Lindsay, R. C. L. (2007). Radical alternatives to traditional lineups. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness

Edmonds v. Mississippi, 955 So.2d 787 (2007).


doi:10.1177/0956797611422918


doi:10.1023/A:1022330011811


doi:10.1023/B:LAHU.0000022323.74584.f5


doi:10.1023/A:1024814009769


doi:10.1016/j.forsciint.2006.02.039


doi:10.1037/a0030609


doi:10.1177/0146167297236003

doi:10.1177/0956797613481812


183


