6-2014

The Influence of Naive and Media-Informed Beliefs on Juror Evaluations of Forensic Science Evidence

Victoria Zoe Lawson

Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Psychology Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/242

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
THE INFLUENCE OF NAÏVE AND MEDIA-INFORMED BELIEFS ON JUROR EVALUATIONS OF FORENSIC SCIENCE EVIDENCE

by

VICTORIA ZOE LAWSON

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

Deryn M. Strange

Date

Chair of Examining Committee

Maureen O’Connor

Date

Executive Officer

Julie Blackman

Saul Kassin

Maureen O’Connor

Emily West

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

THE INFLUENCE OF NAÏVE AND MEDIA-INFORMED BELIEFS ON JUROR EVALUATIONS OF FORENSIC SCIENCE EVIDENCE

by

VICTORIA ZOE LAWSON

Advisor: Professor Deryn Strange

The National Academy of Sciences (2009) concluded that with the exception of nuclear DNA, none of the forensic sciences has been scientifically validated. It is not clear, however, that people are aware of these deficiencies. Indeed, people tend to think quite highly of forensic science, and find it to be convincing trial evidence. It is not clear to what extent their erroneous beliefs about validity influence the weight given to such evidence, or how best to challenge these beliefs. In the present research, I examined people’s beliefs about forensic science and how their beliefs influenced their evaluations of forensic evidence. I also investigated the most effective ways to challenge their beliefs either during the trial (i.e., via cross-examination) or prior to the trial (i.e., via the media).

In the first part of the project (Study 1), I investigated pretrial reliability beliefs, and the influence of DNA, fingerprint, toolmark, and bitemark evidence in a homicide trial. The evidence matched or did not match the defendant and was countered by non-substantive, expert-focused, or evidence-focused cross-examination. Forensic evidence was viewed as more reliable than non-forensic evidence, and reliability beliefs influenced people’s perceptions of the evidence. Although participants had some awareness of the comparative reliability of different disciplines, they tended to give too much weight to less valid disciplines. Further, evidence that matched the defendant was viewed as higher quality than evidence that did not match. Although
cross-examination made people more skeptical of the forensic evidence, it did not reduce guilty verdicts.

In the second part of the project, I investigated the effectiveness of fact-based (Study 2) or story-based (Study 3) media reports in challenging people’s beliefs about the validity of bitemark evidence, and whether reading such reports could help them to evaluate forensic evidence more appropriately. I found that a report which used complex language and attacked bitemark evidence from several angles was the most effective fact-based report. An illustrative story by itself was ineffective, but when the story was supplemented with factual information, it appeared to be the more effective than facts presented alone. Possible implications of these findings are discussed.
Acknowledgements

This research was generously supported by grants from the American-Psychology Law Society and the Forensic Psychology Research Institute.

First, thank you to my committee for your comments and guidance in designing the studies and on the final document. To say that I have benefitted greatly from your collective expertise would be an understatement. To Deryn, thank you for being an incredible advisor and friend and for swooping in to rescue me when I was a bit lost at sea. I consider myself lucky to have had you as a mentor and I look forward to many future collaborations (if you thought you were rid of me, you were most definitely wrong). To Saul, thank you for making me a part of your lab. I have learned so much from you that I am almost ready to forgive you for my painful lesson in field data collection; next time, let’s just herd cats—it would be considerably easier. To Maureen, thank you for your patience and for encouraging me to take the time to think things over and consider other viewpoints. I so appreciate your willingness to lend an ear not only about our studies, but about teaching and life in general. To Julie, thank you for graciously agreeing to return to academia to serve on my committee—your insights were much appreciated. And to Emily, I am so glad I got to continue to learn from you even when you were no longer my boss! Your perspective was incredibly helpful in both designing materials and thinking about my results in new and different ways.

Thank you also to my friends inside and outside the program who helped me to survive the last five years. You know who you are, and hopefully also how much I appreciate your support. Thank you especially to Daisy, Brian and Jeff, without whom being in the program would have been a far more painful experience. Daisy, I don’t know what I will do if you leave me for the other side of the country.

Last, but certainly not least, thank you to my family. To my parents, you have been there for me for much longer than the time it took to get this degree. Thank you for a lifetime of support; I know that I owe a large part of getting here to you. To Nicholas, thank you for talking things through with me and for always having faith that I would end up on top. I have missed having you closer on the road to doctor-hood. And to Dave, thank you for your unwavering love and encouragement even when I was at my most difficult to love or encourage. My future is brighter because you are in it. This dissertation is dedicated to you.
# TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION ................................................................. 1

CHAPTER 2: THE VALIDITY OF THE FORENSIC SCIENCES .................................. 4  
  The Validity of the Forensic Sciences .......................................................... 4  
  Potential for bias in forensic identification judgments ................................... 6  
  The Validity of Specific Disciplines ............................................................. 8  
  DNA analysis .................................................................................................. 8  
  Fingerprint analysis ...................................................................................... 9  
  Toolmark analysis ......................................................................................... 9  
  Bitemark analysis ......................................................................................... 10

CHAPTER 3: PART I: THE INFLUENCE OF NAÏVE BELIEFS ................................. 12  
  Naïve Beliefs about Science and Forensic Science ....................................... 16  
  Beliefs about science ................................................................................... 16  
  Beliefs about forensic science ...................................................................... 17

CHAPTER 4: JUROR EVALUATIONS OF SCIENTIFIC AND FORENSIC EVIDENCE ..... 20  
  Factors Influencing Juror Evaluations of Forensic Science Evidence ............. 23  
  Identification versus non-identification ....................................................... 23  
  Form of testimony .......................................................................................... 23  
  Diagnosticity .................................................................................................. 27  
  Individual differences: Need for Cognition .................................................. 29

CHAPTER 5: THE EFFECTIVENESS OF CROSS-EXAMINATION .......................... 34

CHAPTER 6: STUDY ONE OVERVIEW ................................................................. 37

CHAPTER 7: STUDY ONE METHOD .................................................................... 38  
  Participants .................................................................................................... 38  
  Design ............................................................................................................ 39  
  Materials ....................................................................................................... 39  
  Attitudes towards science ............................................................................ 39  
  Post-trial naïve beliefs about forensic science ............................................. 39  
  Transcripts ..................................................................................................... 40  
  Need for Cognition Scale ............................................................................. 42  
  Manipulation checks ..................................................................................... 42  
  Dependent Measures .................................................................................... 43  
  Procedure ....................................................................................................... 44  
  Hypotheses .................................................................................................... 44

CHAPTER 8: STUDY ONE RESULTS .................................................................... 46  
  Hypothesis 1 ................................................................................................... 46  
  Hypothesis 2 ................................................................................................... 47
Hypothesis 3.............................................................................................................48
  Verdicts .................................................................................................................48
  Verdict confidence .................................................................................................49
  Murder probability estimates ...............................................................................49
  Source probability estimates ..............................................................................50
Hypothesis 4.............................................................................................................51
  Perceptions of expert credibility ...........................................................................52
  Perceptions of evidence quality ...........................................................................52
  Perceived understanding of the expert and evidence .........................................53
Hypothesis 5.............................................................................................................54
Hypothesis 6.............................................................................................................54
  Verdicts and verdict confidence ...........................................................................55
  Probability estimates .............................................................................................56
  Perceptions of expert credibility ...........................................................................56
  Perceptions of evidence quality ...........................................................................57
  Perceived understanding of the expert and evidence .........................................57
Hypothesis 7.............................................................................................................58
  Specific and general naïve beliefs .......................................................................59
  Verdicts ...................................................................................................................59
  Verdict confidence ................................................................................................60
  Probability estimates .............................................................................................61
  Perceptions of expert credibility ...........................................................................62
  Perceptions of evidence quality ...........................................................................63
Hypothesis 8.............................................................................................................64
  Verdicts and verdict confidence ...........................................................................64
  Probability estimates .............................................................................................65
  Perceptions of expert credibility ...........................................................................66
  Perceptions of evidence quality ...........................................................................67
  Perceived understanding of the expert and evidence .........................................67

CHAPTER 9: STUDY ONE DISCUSSION .....................................................................69

CHAPTER 10: PART II: THE INFLUENCE OF MEDIA-INFORMED BELIEFS ..........80
  Pretribal Publicity ..................................................................................................80
  Correcting Misinformation .....................................................................................83

CHAPTER 11: STUDIES TWO AND THREE OVERVIEW ......................................88

CHAPTER 12: STUDY TWO METHOD .....................................................................90
  Participants .............................................................................................................90
  Design .....................................................................................................................90
  Materials ................................................................................................................91
  Articles ....................................................................................................................91
    Control articles .....................................................................................................91
    Target articles ......................................................................................................91
    Pilot testing ..........................................................................................................92
Beliefs about the reliability of forensic sciences questionnaire ........................................... 94
Transcripts .......................................................................................................................... 95
Need for Cognition Scale ................................................................................................. 95
Manipulation checks ......................................................................................................... 95
Dependent Measures ......................................................................................................... 96
Procedure ........................................................................................................................... 96
Hypotheses ......................................................................................................................... 97

CHAPTER 13: STUDY TWO RESULTS .................................................................................. 99
Hypothesis 9 ......................................................................................................................... 99
Hypothesis 10 ....................................................................................................................... 100
Hypothesis 11 ....................................................................................................................... 100
  Perceptions of guilt .......................................................................................................... 101
  Perceptions of expert credibility ..................................................................................... 102
  Perceptions of evidence quality ..................................................................................... 102
  Perceived understanding of the expert and evidence ..................................................... 104
Hypothesis 12 ....................................................................................................................... 104
  Pretrial validity beliefs ................................................................................................. 105
  Perceptions of guilt ......................................................................................................... 105
  Perceptions of the expert and evidence ....................................................................... 106

CHAPTER 14: STUDY TWO DISCUSSION ......................................................................... 108

CHAPTER 15: STUDY THREE METHOD ........................................................................... 114
  Participants ....................................................................................................................... 114
  Design ............................................................................................................................... 114
  Materials .......................................................................................................................... 114
    Articles ......................................................................................................................... 114
      Control articles .......................................................................................................... 114
      Target articles .......................................................................................................... 115
    Pilot testing .................................................................................................................... 115
  Beliefs about the reliability of forensic sciences questionnaire ........................................ 116
  Transcripts ....................................................................................................................... 116
  Need for Cognition Scale ............................................................................................. 116
  Manipulation checks ...................................................................................................... 116
Dependent Measures ......................................................................................................... 117
Procedure ........................................................................................................................... 117
Hypotheses ......................................................................................................................... 117

CHAPTER 16: STUDY THREE RESULTS ........................................................................... 118
Hypothesis 13 ..................................................................................................................... 118
Hypothesis 14 ..................................................................................................................... 118
  Pretrial validity beliefs ................................................................................................. 119
  Perceptions of guilt ........................................................................................................ 120
  Perceptions of expert credibility ................................................................................. 121
  Perceptions of evidence quality ................................................................................... 121
LIST OF TABLES

Table 1: Comparisons between jury and student samples on mean ratings for the accuracy and persuasiveness of different evidence types ................................................................. 140

Table 2: Study 1. Percentage (n) of participants agreeing with individual items comprising beliefs-in-science and reservations-about-science composite variables ........................................... 141

Table 3: Study 1. Reliability ratings for each type of evidence on a 1-5 point scale and results of paired-sample t-tests for each evidence comparison and number (percentage) of participants saying they have never heard of the type of evidence in lieu of rating .......................................................... 142

Table 4: Study 1. Percentage (n) of guilty verdicts in different evidence type conditions dependent on whether the expert testified that the defendant was the source of the evidence (match) or not the source of the evidence (non-match) ................................................................. 143

Table 5: Study 1. Mean (SD) verdict confidence (1-7 scale) in different evidence conditions and when the evidence matched versus did not match the defendant ......................................................... 144

Table 6: Study 1. Mean (SD) estimates of the probability that the defendant murdered the victim and that he was the source of the evidence when the evidence matched and did not match the defendant dependent on type of evidence and type of cross-examination.............................. 145

Table 7: Study 1. Mean (SD) ratings of the credibility of the expert (1-7) and of the validity and helpfulness of the evidence (1-6) in addition to participants’ perceptions of how well they understood the evidence (1-7) dependent on the type of evidence, whether the evidence matched the defendant, and the type of cross-examination used ......................................................... 146

Table 8: Study 1. Mean (SD) pretrial beliefs-in-science (4-16) and reservations-about-science ratings (3-12) for low, moderate, and high NC participants ................................................................. 147

Table 9: Study 1. Mean (SD) confidence in verdict choice (1-7), and ratings of the credibility of the expert (1-7), the validity and helpfulness of the evidence (1-6), and perceived understanding of the expert’s testimony and his forensic technique (1-7) for low, moderate and high NC participants ......................................................................................................................... 148

Table 10: Study 1. Mean (SD) estimates of the probability that the defendant was the source of the evidence for low, moderate, and high NC participants dependent on whether the evidence matched or did not match the defendant ................................................................................................................................. 149

Table 11: Study 2. Validity-belief ratings for control participants only for each type of evidence and results of paired-sample t-tests for each evidence comparison and average knowledge ratings for each evidence type ......................................................................................................................... 150

Table 12: Study 2. Mean (SD) validity-belief ratings in the different article conditions for bite mark and other evidence types ......................................................................................................................... 151
Table 13: Study 2. Mean (SD) ratings of the probability that the defendant was the source of the evidence (1-100), the credibility of the expert (1-9), and the validity and helpfulness of the evidence (1-6) dependent on which article was read .................................................................152

Table 14: Study 2. Mean (SD) ratings of the probability that the defendant was the source of the evidence dependent on which article was read for low, moderate, and high NC participants ....153

Table 15: Study 2. Mean (SD) ratings of perceived understanding of the expert’s testimony and his forensic technique (1-9) for low, moderate, and high NC participants ........................................154

Table 16: Study 3. Validity-belief ratings for control participants only for each type of evidence and results of paired-sample t-tests for each evidence comparison and average knowledge ratings for each evidence type .................................................................155

Table 17: Study 3. Mean (SD) validity-belief ratings in the different article conditions for bite mark and other evidence types .................................................................156

Table 18: Study 3. Mean (SD) ratings of the probability that the defendant was the source of the bitemark or toolmark evidence dependent on which article was read ........................................157

Table 19: Study 3. Mean (SD) ratings of the credibility of the expert (1-9), and perceived understanding of the expert’s testimony and his forensic technique (1-9) dependent on which article was read ........................................................................................................158

Table 20: Study 3. Mean (SD) ratings of the validity of the evidence (1-6) overall and in bitemark and toolmark conditions dependent on which article was read ........................................159

Table 21: Study 3. Mean (SD) ratings of perceived understanding of the expert’s testimony and his forensic technique (1-9) for low, moderate, and high NC participants.................................160
CHAPTER 1: GENERAL INTRODUCTION

In 1985, Robert Lee Stinson was convicted of the murder of 63-year-old Ione Cychosz. Although Stinson lived near the lot where the victim was found, the only evidence directly linking him to the crime was the testimony of two experts that Stinson’s teeth matched several bitemarks found on the victim’s body (State v. Stinson, 1986; www.innocenceproject.org). Dr. Lowell Thomas Johnson made models of Stinson’s teeth and photos of his dentition and of the bitemarks, and testified regarding his comparison of Stinson’s dentition to the bitemarks. He concluded that the bitemarks were identical to Stinson’s teeth, and claimed there was a zero margin of error. Dr. Raymond Rawson, the chairman of the Bitemark Standards Committee of the American Board of Forensic Odontologists, also testified regarding the high quality of the evidence, calling it “overwhelming.” He concluded that the bitemarks had been made by Stinson to a reasonable degree of scientific certainty. The jury found Stinson guilty of first-degree murder and he was sentenced to life in prison.

The bitemark evidence constituted the only direct evidence against Stinson. Thus, to find Stinson guilty the jurors must have believed that an expert’s testimony regarding a bitemark match was sufficient evidence to support a verdict of guilt. Their faith in the evidence, however, was misplaced. Bitemark analysis has been roundly criticized in recent years, with serious doubt cast on the assertion that any scientific basis exists for the practice (Harris, 2012; National Academy of Sciences, 2009). In fact, even one of the most (in)famous bitemark analysts has since rejected the practice (Mitchell, 2012). Thus in Stinson’s case, the jurors’ erroneous beliefs regarding the validity of bitemark evidence and the probative value of such testimony sent an innocent man to jail for over two decades; it was not until 2009, after DNA proved his innocence, that Stinson was released (www.innocenceproject.org).
Stinson’s case is not an anomaly: improper or unreliable forensic sciences were involved in more than 50% of the 316 cases on the Innocence Project’s current list of DNA exonerations (www.innocenceproject.org). Further, false or misleading evidence was known to have been involved in 22% of the exonerations tracked to date by the National Registry of Exonerations (www.law.umich.edu/special/exoneration/Pages/about.aspx), which includes both DNA and non-DNA exonerations. At present, however, we do not know how many other people have been wrongfully convicted due to problematic forensic science, particularly given that many states have historically not had regulations in place to preserve the evidence that would allow for future testing (www.innocenceproject.org). As such, the known wrongful conviction cases provide concrete examples of what is likely a much larger problem: the belief that the forensic sciences are valid—held by the justice system and regularly used to support admitting evidence against those charged with crimes (Cole, 2010)—may be erroneous and may contribute to a significant number of wrongful convictions. Therefore, if future wrongful convictions are to be prevented, the ability of jurors to appropriately evaluate forensic science is an issue of great importance. However, the extent to which these evaluations may be affected by prior beliefs regarding the validity of forensic science has not yet been established. Moreover, if jurors do have erroneous beliefs regarding the validity of forensic science, the best way to counter those beliefs is not yet clear.

Jurors’ evaluations of forensic science evidence may be influenced by their general beliefs about science and by prior beliefs about the soundness of the particular type of science presented at trial. Therefore, challenging naïve beliefs about forensic science may have a direct impact on jurors’ perceptions of the evidence and their subsequent judgments. Additionally, challenging naïve beliefs may improve jurors’ ability to weigh evidence accurately. In a series of
studies, I examined people’s beliefs about forensic science evidence and how those beliefs influenced their evaluations of trial evidence and subsequent decision-making. I also investigated the most effective ways to challenge these beliefs either during the trial (i.e., via cross-examination) or prior to the trial (i.e., via the media). In the first part of the current project, I examined mock jurors’ responses to different types of forensic science evidence. I also examined the extent to which their perceptions of the evidence may have been influenced by their prior naïve beliefs regarding the specific science presented and science more broadly. Further, I examined whether the traditional legal safeguard of cross-examination could effectively challenge forensic science testimony. Here I was particularly interested in whether cross-examination could counter jurors’ preconceived notions about the extent to which forensic evidence was a sufficient basis for a verdict decision. In the second part of the project, I moved outside of the trial process itself and investigated whether educating jurors about the problems with forensic science using different types of factual or narrative-based media reports could reduce jurors’ beliefs in the validity of forensic evidence and improve their ability to weigh such evidence appropriately. Taken together, these studies provided insight into a major challenge facing the legal system—how to best and most effectively help ensure that the forensic sciences are not given undue deference at trial.
CHAPTER 2: THE VALIDITY OF THE FORENSIC SCIENCES

What should people believe about the validity of the forensic sciences? Is faith in forensic science generally misplaced? According to the scientific community, the answer for most disciplines is yes.

The Validity of the Forensic Sciences

Following an extensive review of the field, the National Academy of Sciences (NAS) released a report on the forensic sciences in 2009 (NAS Report). They concluded that most of the forensic sciences have not yet been sufficiently validated. The report noted that practitioners in many forensic science disciplines draw conclusions in terms of matching a particular source to the exclusion of others, but found that such conclusions are rarely warranted. Indeed, the authors concluded that “with the exception of nuclear DNA analysis … no forensic method 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). Many of the forensic disciplines are lacking the kind of research that would be required to determine where, and to what degree of precision, conclusions can be drawn. Moreover, they lack the kind of research that would determine how to avoid the potential for human error, particularly error due to bias. The authors of the report also noted that there is wide variation in the extent to which protocols exist to guide the analyses and conclusions drawn by experts in the different disciplines. In short, the forensic sciences vary quite widely in the extent to which they can be considered valid and reliable sciences.

Indeed, there is variation in the presence/number of validation studies, the amount of regulation regarding methods of analysis, the potential for results to be affected by bias or other sources of error, and the diagnosticity of the information provided by a match (i.e., given that a
match has been found, how likely it is that the purported source is actually the source of the evidence; Kaasa, Peterson, Morris, & Thompson, 2007; Koehler, 2011). One basic distinction pointed to by the NAS is that some disciplines are based in the laboratory, while others involve the interpretation of observed patterns by an expert. The former generally have a greater scientific basis and include such disciplines as nuclear DNA, mitochondrial DNA, toxicology, and drug analysis. The latter disciplines vary more widely in their scientific basis and include such disciplines as the analysis and comparison of fingerprints, handwriting samples, toolmarks, hair samples, and bitemarks.

The NAS Report also noted several problems with the forensic sciences that, in part, stem from their role as part of a criminal investigation including the impetus for their development, the goal of individualization, and the potential for bias. For example, many of the forensic sciences were developed in crime labs rather than in scientific laboratories. Their goal is to aid in criminal investigations; thus, historically, research has not been a priority nor has there been much funding available for validation research. Additionally, many of the forensic sciences have individualization as the goal of their analyses. That is, they are interested in determining whether a specific source of the evidence has been identified “to the exclusion of all other possible sources” (p. 44). In part this goal is related to the aim of practitioners to aid in criminal investigation, which often involves the building of a case against a specific person. However, the NAS Report explained that such determinations cannot be made simply by comparing the evidence to a particular source. Instead, such determinations require knowledge of the frequency in the population of both variations in attributes (e.g., a particular marking) and of the likelihood that two sources in the population could share the same attributes. Many of the forensic sciences have not undertaken such studies and no population statistics underlie their conclusions. The
NAS Report also pointed to the potential for bias as a problem for many of the sciences: the disciplines tending to lean more toward subjective judgments rather than more objective laboratory analysis are considerably more susceptible to bias (see also Thompson & Cole, 2007).

**Potential for bias in forensic identification judgments.** The NAS Report noted that a number of human biases may influence forensic judgments; however, many of these may result from, or be related to, forensic confirmation bias (Kassin, Dror, & Kukucka, 2013). We know from basic research on cognitive biases that expectations can influence outcomes and perceptions (Darley & Gross, 1983; Lord, Ross, & Lepper, 1979; Nickerson, 1998; Saks, Risinger, Rosenthal, & Thompson, 2003). Thus, if an analyst expects his or her analysis to have a certain outcome (e.g., that it will match the police suspect), their expectation may influence their analysis and, ultimately, their conclusion (Kassin et al., 2013; NAS, 2009; Saks et al., 2003). A series of studies conducted on fingerprint analysis by Dror and Charlton (Dror & Charlton, 2006; Dror, Charlton, & Peron, 2006; Dror, Peron, Hind, & Charlton, 2005) are an excellent demonstration of the potential for error in forensic evaluations due to bias, particularly as fingerprint examinations are generally viewed as highly reliable and valid by both laypeople and experts (Dror et al., 2005; Garrett & Mitchell, 2013; Lieberman, Carrell, Miethe, & Krauss, 2008; NAS, 2009).

Dror and Charlton’s (2006; Dror et al., 2005, 2006) studies examined the influence of context on fingerprint analysis. The first study (Dror et al., 2005) was conducted with university students. Dror et al. found that students were more likely to conclude that ambiguous pairs of prints matched when they were presented with highly emotional case information than when they were presented with more neutral information. Experienced fingerprint experts fared no better. In the next two studies (Dror & Charlton, 2006; Dror et al., 2006), experts were provided with pairs
of prints that they had previously examined during the course of their regular case work. These prints were presented to them by colleagues, thus they were unaware that these particular examinations were part of the study. In the first study (Dror et al., 2006), five experienced experts were each presented with a pair of prints that they had previously concluded matched. However, they were told that the prints had *erroneously* been called a match by other examiners in a high profile case. Although told to ignore the contextual information, all but one of the examiners concluded that the prints they examined did not match, contradicting their prior judgment.

In the next study (Dror & Charlton, 2006), six experienced experts were each presented with pairs of prints that they had previously determined to match or not to match. For half of the pairs, examiners were given information suggesting an identification or a non-identification, while no contextual information was given for the other half. Across all examinations, experts changed opinions in 12% of the cases, while only two of the six examiners made decisions totally consistent with their prior conclusions. Examiners were more likely to change their decision when contextual information was provided than when it was not. We can assume that the biasing effect of context is not limited to fingerprint analysis and will affect any forensic examination which involves the subjective determination of a match between evidence and a particular source (Dror & Rosenfeld, 2008; Kassin et al., 2013; Saks et al., 2003; but see Kerstholt, Paashuis, & Sjerps, 2007; Kerstholt et al., 2010). Indeed, similar results have been found in other disciplines including handwriting analysis (Kukucka & Kassin, 2013; L.S. Miller, 1984), microscopic hair examination (L.S. Miller, 1987), polygraph (Elaad, Ginton, & Ben-Shakhar, 1994), forensic anthropology (Nakhaeizadeh, Dror, & Morgan, 2013), and auditory evidence interpretation (Lange, Thomas, Dana, & Dawes, 2011).
The Validity of Specific Disciplines

Regarding specific types of evidence, the NAS Report gave an overview of the current state of the science on a number of different techniques. The four that I focus on in this project are DNA, fingerprint, toolmark, and bitemark analysis. These disciplines represent the full range of validity: DNA was held up as the standard against which the other sciences are to be judged, while the authors pointed to both toolmark and bitemark analysis as examples of types of tests that “have never been exposed to stringent scientific scrutiny” (p. 42).

**DNA analysis.** DNA is considered the gold standard for forensic evidence because extensive validation testing has been conducted on DNA analysis, it relies less on subjective judgment than most other types of forensic science, and population statistics have been compiled which provide information about the likelihood and probative value of a match (NAS, 2009). In addition to the above, the NAS Report pointed to five specific factors that make DNA analysis scientifically sound:

1) there are biological explanations for individual-specific findings; 2) the 13 STR loci used to compare DNA samples were selected so that the chance of two different people matching on all of them would be extremely small; 3) the probabilities of false positives have been explored and quantified in some settings (even if only approximately); 4) the laboratory procedures are well specified and subject to validation and proficiency testing; and 5) there are clear and repeatable standards for analysis, interpretation, and reporting.

(p. 133)

It should be noted, however, that even DNA is not 100% accurate: contamination is a particular problem, and given that DNA analyses are conducted by humans, there is always the potential for error (Lieberman et al., 2008). Further, interpreting complicated mixtures may involve more
subjective judgment than simple single-source identifications, which makes such conclusions susceptible to the same types of bias as other types of forensic science (Dror & Hampikian, 2011).

**Fingerprint analysis.** Although there are Automated Fingerprint Identification Systems that ostensibly allow for the automatic identification of fingerprints contained in criminal databases, the NAS Report pointed out that subjective expert interpretation forms the basis for the majority of fingerprint identifications. Latent prints found at a crime scene are often distorted or degraded in some way; thus the expert must use subjective judgment to determine whether a small difference between features on the latent and a suspect’s print are due to distortion in the latent or the fact that they do not match. Partially for that reason, no population statistics have been developed and little research has been devoted to the creation of population statistics. Further, the techniques relied upon by fingerprint experts have not yet been validated, and error rates have not yet been established; as demonstrated in Dror and Charlton’s studies (Dror & Charlton, 2006; Dror et al., 2006), there is also considerable potential for error due to bias. However, the NAS Report also noted several positive points about fingerprint analysis: that fingerprint examiners have at least developed protocols for how fingerprint examinations should be carried out, that research on fingerprint analysis is building, and that “the amount of details available in friction ridges … [makes it at least] plausible that a careful comparison of two impressions can accurately discern whether or not they had a common source” (p. 142).

**Toolmark analysis.** The NAS Report noted that toolmark analysis suffers from many of the same problems as fingerprint analysis. That is, analysis relies heavily on the subjective judgment of the examiner, no population statistics exist, and error rates are not currently known. The authors did note that some research has been conducted on toolmark analysis and that a
sufficient basis *might* exist for such analyses to narrow the field of possible tools that left a mark. However, they cautioned that while it is theoretically possible that individual patterns might be able to suggest a specific source of a particular toolmark, not enough research exists in order to make such a determination. Further, no protocols to guide toolmark examination have been established and it is not clear what degree of correspondence between a questioned mark and a particular tool is necessary in order for them to be declared a match.

**Bitemark analysis.** The NAS Report explained that there are several areas of forensic odontology—the legal application of dentistry—including not only bitemark comparison but also interpreting oral injuries, dental malpractice, and identifying unknown remains. Bitemark analysis is used most often in criminal cases and is also the most controversial. Bitemark analysis relies almost completely on the subjective judgment of the expert and as such is susceptible to many of the same problems and biases as fingerprint and toolmark analysis, including the lack of population statistics. The NAS Report noted that bitemark experts often vary quite widely in their analyses of the same evidence, and that the potential for error due to bias may be greater in bitemark analysis than in other types of analyses due to the often highly sensational cases in which they tend to be used (see also Page, Taylor, & Blenkin, 2012). Although the NAS Report noted that it is possible that bitemark analysis may be used to exclude suspects, the authors concluded that no scientific basis exists to determine that a bitemark matches a particular suspect, or that such a determination is even possible. The report pointed to three particular problems with bitemark analysis:

1) The uniqueness of the human dentition has not been scientifically established. 2) The ability of the dentition, if unique, to transfer a unique pattern to human skin and the ability of the skin to maintain that uniqueness has not been scientifically established. …
3) A standard for the type, quality, and number of individual characteristics required to indicate that a bitemark has reached a threshold of evidentiary value has not been established. (pp. 175-6)

Indeed, so little research has been conducted that the NAS Report concluded that it is not possible at present to determine whether the methods and conclusions of bitemark analysts have any probative value. Thus, it is unclear whether a bitemark match should be considered probative evidence at trial, and certainly it should not be considered sufficient evidence to support a conviction.
CHAPTER 3: PART I: THE INFLUENCE OF NAÏVE BELIEFS

The conclusions from the NAS Report indicate that if people believe that the forensic sciences are scientifically valid, with the exception of DNA, their beliefs are erroneous. To what extent, then, might these naïve beliefs regarding forensic science influence their perceptions of trial evidence and their ultimate judgments if they are called as jurors in criminal cases? In general, we know that people’s personal beliefs and attitudes may color their perceptions of the information they receive from the world, and may also affect the manner in which they evaluate and interpret information even at the expense of accurate appraisal (e.g., Fishbein & Ajzen, 1972; Lord et al., 1979; Nickerson, 1998; Saks et al., 2003). Indeed, given that we tend to interpret new information through the lens of what we already know to be true, prior knowledge and beliefs will necessarily have some influence on the processing of new information (Lombrozo, 2006; Posner, Strike, Hewson, & Gertzog, 1982). In the legal arena, despite the fact that jurors are instructed to focus solely on the evidence presented and any legal guidelines, jurors are not always able to set aside extra-legal information such as their prior beliefs (Daftary-Kapur, Dumas, & Penrod, 2010; Kassin & Sommers, 1997). Indeed, we know jurors do rely on other types of extra-legal factors including such factors as the race of the defendant (Sommers & Ellsworth, 2000), the defendant’s prior criminal record (Greene & Dodge, 1995), and evidence ruled inadmissible (Kassin & Sommers, 1997; Lieberman & Arndt, 2000; Steblay, Hosch, Culhane, & McWethy, 2006). Particularly given that jurors may have strongly held convictions regarding science, such research suggests that they may have similar difficulty in setting aside their pretrial beliefs.

Prior beliefs may be more likely to influence juror decision-making in some circumstances than others; thus it is not entirely clear when and to what extent jurors’
evaluations of forensic science evidence are based on their prior naïve beliefs about the validity of the specific science presented as evidence. One possibility is that pretrial beliefs may act as a heuristic, a type of mental shortcut used in decision-making (Tversky & Kahneman, 1974). Research confirms that people do rely on heuristics, not only in their general decision-making, but also in their decision-making as jurors. Indeed, research suggests that jurors rely on heuristics such as availability (e.g., Brekke & Borgida, 1988; Diamond & Stalans, 1989; Greene & Wade, 1988), and anchoring and adjustment (e.g., Chapman & Bornstein, 1996; Marti & Wissler, 2000) in arriving at their trial judgments. Further, dual-processing models of persuasion (Chaiken, 1980; Petty & Cacioppo, 1986) suggest that jurors may be more likely to rely on heuristics when the trial information is difficult to understand.

In essence, trials are a form of persuasion, wherein the opposing sides attempt to persuade jurors to make a particular decision; thus researchers have examined whether dual-processing models of persuasion might help to explain juror decision-making (see Groscup & Tallon, 2011; McAuliff, Ellis, & Phillips, 2011). Put simply, dual-processing models posit that information may be processed either on an automatic (superficial) level or a deeper, more effortful and deliberative level (Chaiken, 1980; Petty & Cacioppo, 1986). When both motivation and ability are high, information is processed more deeply and systematically; when either is low, information is processed more superficially and people are more likely to use heuristics and other mental shortcuts. Thus, any factor that reduces jurors’ motivation or ability to process information may make them more likely to rely on heuristics, which may include prior beliefs (Groscup & Tallon, 2011; McAuliff et al., 2011).

If prior beliefs function as a heuristic, jurors may be more likely to rely on pretrial beliefs when presented with complex scientific evidence. Indeed, Cooper, Bennett, and Sukel (1996)
demonstrated that complexity could induce jurors to rely on heuristics. They presented mock jurors with trial materials that included expert testimony that was either complex (e.g., used technical scientific jargon) or simple. They also manipulated the quality of the expert, presenting him as either highly credentialed (e.g., degrees from prestigious universities, numerous publications) or moderately credentialed (e.g., degrees from obscure universities, few publications). Thus expert credentials was operationalized as a heuristic cue. Cooper et al. found that mock jurors were more likely to rely on the credentials of the expert in judging the case when the testimony was complex than when it was simple. Jurors may also be more likely to rely on other types of heuristics, include their a priori beliefs regarding expert credibility or scientific validity, when presented with complex, scientific evidence.

If prior beliefs do function as a heuristic, anything that reduces jurors’ ability or motivation to process evidence may make them more likely to rely on their pretrial beliefs (Giner-Sorolla, Chaiken, & Lutz, 2002). For example, research has found that mock jurors are more likely to rely on stereotypes or on prior beliefs about sex discrimination when cognitive load is increased by reducing the amount of time allotted for a task (Giner-Sorolla et al., 2002; Gordon & Anderson, 1995; van Knippenberg, Dijksterhuis, & Vermeulen, 1999). Additionally, information that is consistent with prior beliefs may be more likely to be automatically accepted because it is easier to understand than information inconsistent with prior beliefs. Indeed, there is evidence that people’s initial automatic response may be to believe information that they are able to understand—that the process of comprehending information in itself tends to inspire acceptance—while disbelieving information is a secondary process that requires additional effort (Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993).
The influence of prior beliefs need not be conscious, nor would the purported effect of prior beliefs necessarily be reflected in jurors’ reports of what did influence their evaluations of the evidence. Much of our information-processing occurs unconsciously (Bargh & Chartrand, 1999; Greenwald & Banaji, 1995), and we may be unaware not only about what influences our perceptions and behaviors, but by what mechanisms outside factors influence us (Nisbett & Wilson, 1977). Further, self-report of the influences on decision-making is not necessarily an accurate reflection of the degree to which a particular type of stimulus or information affects our responses (Wilson & Nisbett, 1978). Put simply, we are not the best judges of what does and does not affect us.

Prior beliefs may also have a greater influence on behavior when there is a greater correspondence between the particular belief and the behavior upon which it acts (Ajzen & Fishbein, 1977). If so, beliefs and attitudes that more closely correspond to the decision to be made may be more likely to influence juror decision-making (Giner-Sorolla et al., 2002). Indeed, while attempts to link general characteristics (e.g., demographics) to legal decision-making have met with limited success (Crocker & Kovera, 2011), research suggests that attitudes closely related to the legal system, or to the particular type of evaluation to be made, do affect jurors. For example, strong beliefs in the justice system and reliance on authority may be associated with a greater likelihood of conviction (Kassin & Wrightsman, 1983; Narby, Cutler, & Moran, 1993), internal locus of control may be associated with greater attributions of criminal responsibility (Phares & Wilson, 1972), and attitudes towards psychiatrists and towards the insanity defense may be associated with verdict choices in insanity defense cases (Cutler, Moran, & Narby, 1992).
Naïve Beliefs about Science and Forensic Science

Beliefs about science. Jurors may hold specific pretrial beliefs regarding both the merits of science in general and the validity of particular types of science. General beliefs about the validity of science may influence the degree of skepticism with which particular types of science are viewed, even in the absence of prior knowledge regarding the science itself. A review of surveys conducted from the 1950s to the 1990s using questions developed by the National Science Board (NSB) indicates that Americans are generally interested in science and believe that it is beneficial (J. D. Miller, 2004). About 80% of Americans said that science improved the world over the 40 year time period, 81-94% thought that it improved people’s lives, and roughly 75-80% believed that it provided more opportunities for future generations. At the same time, Americans also expressed reservations about science, with roughly 40% saying that science changes things too rapidly, and roughly 50% concerned that we rely too greatly on science and not enough on religious values and faith. Beliefs in the benefits of science and reservations about science were shown to be distinct concepts with a moderately strong negative correlation; however, when asked to weigh the two directly, 70-75% of people surveyed in the 1980s and 1990s stated that the benefits of science outweighed the potential risks. NSB reports from 2004 and 2006 found similar levels of agreement regarding beliefs in and reservations about science as those from the earlier time period. Relatively similar numbers were also found with a sample of jurors surveyed using the NSB questions in 2003 (Hans, Kaye, Dann, Farley, & Albertson, 2011). Thus, these perceptions may be relatively stable.

General beliefs about science have not necessarily been linked to verdicts, although they may be associated with perceptions of expert evidence. Bornstein (2004) found that general attitudes towards the scientific process were not associated with verdict choices in a liability
case, although they did influence perceptions of one of two testifying scientific experts. That said, there is evidence that different types of general beliefs may have differential effects: Giner-Sorolla et al. (2002) compared the influence of ideological beliefs—beliefs tied to one’s worldview—to case validity beliefs—preconceived notions about the likely strength of a particular type of case. They found that ideological beliefs had a greater effect, while case validity beliefs affected mock jurors only when they had limited time in which to read the experimental materials. Nevertheless, it is possible that prior beliefs about a specific type of science may influence the persuasiveness of that science when it is presented as trial evidence.

Beliefs about forensic science. Prior research suggests that people do have preconceived notions of the reliability and validity of various types of forensic evidence (Hans et al., 2011; Lieberman et al., 2008). Further, people seem to have a general idea of at least the comparative validity of different types of forensic science; however, it appears that they tend to overestimate the validity of certain types of forensic science. In general, DNA is considered more reliable and valid than other types of evidence (Hans et al., 2011; Lieberman et al., 2008). For example, Hans et al. (2011) asked a sample of jurors to rate the reliability of DNA and several types of non-scientific evidence, and DNA greatly surpassed the reliability ratings of any other type of evidence. Most jurors (95%) rated DNA evidence as either extremely or very reliable, while roughly two thirds of jurors rated police evidence (67%) and expert witness evidence (66%) to be extremely or very reliable, and roughly one third (37%) rated victim evidence to be extremely or very reliable. Eyewitness evidence was viewed as least reliable with only a quarter of jurors (25%) rating it to be extremely or very reliable. However, scientific evidence may generally be viewed more positively than non-scientific evidence, and forensic science may be viewed more positively than other types of science (Lieberman et al., 2008).
Lieberman et al. (2008, Study 1) surveyed both a student and a juror sample regarding their perceptions of the accuracy and persuasiveness of different types of evidence. They included three types of forensic science evidence—DNA, fingerprints, hair/fiber evidence—and several other types of evidence: unspecified expert testimony by scientists, alcohol/drug tests, a suspect’s confession, video surveillance pictures, victim testimony, and eyewitness testimony. Lieberman et al. found that the three types of forensic science received the highest accuracy ratings, with all rated above 88.5% accurate (see Table 1); there were no significant differences between the two samples in accuracy ratings for these three types of evidence. For both jurors and undergraduates, DNA evidence received the highest ratings, followed closely by fingerprints and then hair/fiber evidence. Of the other types of evidence, only ratings for video surveillance pictures were similar. Unspecified expert testimony by scientists received lower ratings than the specific forensic sciences and accuracy ratings were significantly higher for students than they were for jurors. Eyewitness statements received the lowest accuracy ratings.

In addition to asking about accuracy, Lieberman et al. (2008) asked both jurors and students to rate the persuasiveness of a subset of evidence types. DNA was rated as most persuasive, and with the least variability, by both samples. That said, it was viewed as somewhat more persuasive when it was found on the victim than when it was found at the crime scene (see Table 1). Despite lower accuracy ratings, a suspect’s confession was rated as more persuasive than hair/fiber evidence; however, hair/fiber evidence still received high ratings of persuasiveness from both samples. Positive identifications by the victim or an eyewitness were viewed as somewhat—though not greatly—less persuasive.

Lieberman et al. (2008) also provided evidence that pretrial beliefs about a specific type of evidence may influence jurors’ verdicts. In a subsequent experiment (Study 3), they examined
the influence of prior beliefs about DNA evidence on juror judgments in a trial where the primary evidence was DNA. They embedded four questions regarding beliefs in the strength and reliability of DNA evidence in a larger legal attitudes questionnaire, which participants completed prior to the presentation of trial materials. By combining responses to the four questions, Lieberman et al. created a composite pretrial trust in DNA variable. Lieberman et al. found that pretrial trust in DNA significantly predicted verdict choice, in addition to jurors’ ratings of the probability that the defendant was guilty and a composite variable incorporating verdict and verdict confidence. Put another way, jurors with higher opinions of DNA evidence were more likely to vote for conviction and to do so with greater confidence. Moreover, they thought it was more likely that the defendant murdered the victim than jurors with lower opinions of DNA evidence.
CHAPTER 4: JUROR EVALUATIONS OF SCIENTIFIC AND FORENSIC EVIDENCE

While they may be correct about the value of DNA evidence in many circumstances, it is not clear that jurors are able to consider the shortcomings of less valid forms of evidence in their evaluations of expert testimony or in how they weigh the conclusions of scientific experts in reaching verdict decisions. Although jurors may be competent decision-makers under many circumstances (Vidmar, 2005), research has shown that jurors may have difficulty in understanding and evaluating complex scientific testimony and that they may not be able to reliably discriminate between valid and invalid science (e.g., Kovera, McAuliff, & Hebert, 1999; McAuliff & Duckworth, 2010; McAuliff, Kovera, & Nunez, 2009).

For example, Kovera and McAuliff (Kovera et al., 1999; McAuliff et al., 2009) examined whether mock jurors were sensitive to variations in internal validity in a hostile work environment case. In their studies, an expert testified regarding a study that she had conducted on sexual harassment and the researchers manipulated the internal validity of the study. They found that their participants were not sensitive to variations in construct validity (Kovera et al., 1999), or to the presence of a confound, or experimenter bias (McAuliff et al., 2009). Participants did rate the expert as less credible and the evidence as lower in quality when the study was missing a control group than when it was internally valid. However, participants’ liability judgments were ultimately not affected by the missing control group (McAuliff et al., 2009).

Such problems are not limited to civil litigation. McAuliff and Duckworth (2010) had community members read an extended trial summary of a child sexual abuse case including the testimony of a defense expert. The expert’s testimony included a description of one of his recent studies on children’s suggestibility. The researchers presented the study as either valid, missing a control group, including a confound, or susceptible to experimenter bias; they additionally
manipulated whether or not the study was published in a peer-reviewed journal. The internal validity of the study had no influence on participants’ ultimate verdicts. When the study had been published, however, mock jurors thought the defendant was more likely to be guilty when the study was missing a control group than when it was valid. Regardless of publication status, mock jurors rated the study as higher in quality when it was valid than when it was missing a control group. Neither the inclusion of a confound nor experimenter bias influenced perceptions of guilt or the quality of the evidence.

Jurors may also have difficulty with probabilistic or other statistical testimony (e.g., Faigman & Baglioni, 1988; Kaye & Koehler, 2001), including testimony on the random match probabilities (RMPs) underlying determinations made in DNA and other forensic identifications (e.g., Koehler, Chia, & Lindsey, 1995; Schklar & Diamond, 1999; Smith, Penrod, Otto, & Park, 1996; Thompson, Kaasa, & Peterson, 2013). For example, Koehler et al. (1995) presented jury-eligible adults with a murder case in which genetic material from under the victim’s fingernails was found to match the defendant’s DNA profile. They varied whether participants were presented with a laboratory error rate and whether they were presented with a very small RMP. The RMP is the likelihood that the profile found would match a random person in the reference population (e.g., Whites); thus, a very small RMP in the absence of laboratory error would suggest a high probability that the profile did in fact belong to the defendant. However, the error rates included were both much higher than the RMP, essentially rendering it irrelevant information. Koehler et al. found that participants were much more likely to render guilty verdicts when presented with a small RMP, regardless of whether or not they were presented with error rates. This finding suggests that they were overly convinced by the RMP, even when it was irrelevant.
Unsurprisingly, then, research has shown that DNA evidence is convincing to jurors (e.g., Golding, Stewart, Yozwiak, Djadali, & Sanchez, 2000; Lieberman et al., 2008), and may, appropriately, be more convincing than other types of evidence. Lieberman et al. (2008, Study 2) had participants read two-page summaries of a sexual assault case which included one of four forensic evidence types: DNA, blood type, fingerprint, and hair fiber. Participants’ ratings of the likelihood that the defendant committed the crime were higher in the DNA condition (M = 65%) than they were in the fingerprint (M = 48%), or hair fiber (M = 43%) conditions. The DNA and blood type (M = 56%) conditions did not differ; however, the authors’ examination of the manipulation check questions indicated that some participants may have thought that the blood type evidence was in fact DNA evidence. Thus, their results suggest that DNA evidence may be more persuasive than other types of forensic evidence.

However, jurors also find less valid types of forensic science very convincing. Physical evidence (i.e., tangible objects or trace evidence connected to a crime scene) may be viewed as more persuasive than other types of evidence (Skolnick & Shaw, 2001), and even non-DNA forensic sciences appear to be more persuasive than other types of physical evidence (Lieberman et al., 2008). Lieberman et al. (2008, Study 1) presented a student sample with one of several brief case scenarios. The scenarios varied the type of evidence presented (DNA, fingerprints, hair fibers, eyewitness testimony, or victim testimony), whether the evidence was exculpatory or incriminating, and whether the crime that occurred was rape or murder. In the rape scenario, when the evidence was incriminating, the percentage of guilty verdicts was uniformly high (ranging from 92% to 100%) for DNA evidence, hair evidence, fingerprint evidence, and victim testimony; it was significantly lower only for eyewitness testimony (75%). However, when the evidence was exculpatory, there were much lower conviction rates when exculpatory DNA
evidence was presented (15%) than in the other conditions (combined conviction rate of 54%). The opposite pattern was found in the murder scenario: when the evidence was incriminating, there were more guilty verdicts in the DNA condition (100%) than in the other conditions (combined conviction rate of 64%), while when the evidence was exculpatory, there was not a significant difference in verdicts based on evidence type.

Factors Influencing Juror Evaluations of Forensic Evidence

Identification versus non-identification. Lieberman et al.’s (2008) results suggest that evidence presented as matching the suspect and evidence presented as not matching the suspect may not be considered equally probative. McAllister and Bregman (1986) found a similar discrepancy with fingerprint evidence. They found that a fingerprint identification was associated with higher likelihood of guilt and severity of sentence ratings, while a fingerprint non-identification was not significantly different from control. Therefore, while jurors were influenced by forensic evidence implicating the defendant, they failed to take into account potentially exculpatory evidence. Similarly, Garrett and Mitchell’s (2013) participants thought that it was equally likely that a defendant in a robbery trial committed the crime when he was excluded from a latent print found at the scene, when the examiner’s results were inconclusive, and when no fingerprint examiner testified at all. These results suggest that jurors may generally be more influenced by a forensic identification than by a non-identification. Nevertheless, it is not clear whether the extent to which this mismatch occurs generalizes across all types of science or whether certain types of non-identifications are more compelling than others (e.g., DNA evidence; Lieberman et al., 2008).

Form of testimony. The manner in which the expert conveys the forensic identification to jurors may also influence jurors’ perceptions of the evidence. The way that jurors interpret and
weigh probabilistic information such as DNA evidence may depend heavily on how the numbers are presented to them. For example, research suggests that frequencies (e.g., 1 out of 100) are preferred to probabilities (e.g., 1% chance) and small ratios (e.g., 1/10) to large ratios (e.g., 1/1000) (Koeher & Macchi, 2004). However, jurors may prefer not to hear numbers at all: McQuiston-Surrett and Saks (2009) found that a microscopic hair identification expressed qualitatively was generally more persuasive than when it was expressed quantitatively. They compared two forms of qualitative testimony—that the examined hairs “match” or that they are “similar-in-all-microscopic-characteristics” to three forms of probability-based testimony—subjective probability, objective single-probability, or objective multi-frequency probability. Subjective-probability testimony estimated the likelihood of a match but the expert additionally explained that the probability was not based on objective criteria—he testified that he had made a subjective judgment about the probability of a match. Objective single-probability testimony gave the likelihood of a match in terms of the likelihood that a person chosen at random in the population would have a hair matching the one found. Objective multi-frequency probability testimony gave the likelihood of a match in terms of the number of people in the population likely to have hair that matched the sample. Participants were more persuaded by qualitative (“match” or “similar-in-all-microscopic-characteristics”) and objective single-probability opinion testimony than they were by subjective-probability or objective multi-frequency probability testimony.

When qualitative testimony is given, it is also possible that the specific wording used by the expert may influence jurors’ judgments. Using brief summaries of the evidence to be presented in a robbery case, Garrett and Mitchell (2013) examined the influence of different forms of fingerprint testimony on participants’ judgments of the likelihood that the defendant
committed a robbery and that he was the source of a fingerprint found at the scene. In the first category of testimony, the *simple positive match*, the expert testified that he had found a match between the defendant’s print and the print found at the crime scene (e.g., “A latent fingerprint found at the scene was individualized as the left thumb print of the defendant”). In the second category of testimony, the *bolstered positive match*, the expert supplemented the match with a statement of practical or scientific certainty (e.g., “I conclude to a reasonable degree of scientific certainty that the latent fingerprint found at the scene came from the same source as the left thumb print on the ink card labeled as taken from the defendant”) or about the very remote likelihood that the fingerprint could have been made by anyone other than the defendant (e.g., “The latent fingerprint found at the scene was individualized as the left thumb of the defendant. The likelihood the impression was made by a different source is so remote that it is considered to be a practical impossibility”). In the *inconclusive match* category, the expert testified that he could not determine whether the print came from the defendant (“The friction ridge impression did not have sufficient detail to permit a conclusion whether it originated from the defendant or not. The result was inconclusive”), while in the *qualified match* category, the expert either admitted that the print could match another individual (“The latent fingerprint found at the scene was individualized as the left thumb of the defendant. However, it is possible that the print in question could have come from someone else”) or simply said that the defendant could not be excluded as the source of the latent fingerprint. Garrett and Mitchell also examined whether individuals differing in numeracy—a measure of mathematical comfort and ability—weighed fingerprint evidence differently. Participants in the *simple positive match* and *bolstered positive match* conditions rated it significantly more likely that the defendant committed the crime and was the source of the evidence than participants in the *inconclusive* or *qualified match*
conditions. However, the *simple positive match* and *bolstered positive match* conditions did not differ. Numeracy did not influence participants’ ratings of the likelihood that the defendant committed the crime, but more numerate participants expressed greater confidence in their likelihood ratings and rated it more likely that the defendant was the source of the evidence than less numerate participants.

In a follow-up study, Garrett and Mitchell (2013) investigated whether giving participants information about the methodology used in fingerprint analysis and the expert’s discounting or acknowledgement of error rates influenced their perceptions of fingerprint testimony. Either a simple or bolstered match was presented: the expert testified that the “fingerprint found on the gun was individualized as the right thumb of the defendant” or added the information that it was practically impossible that the latent print had been made by someone other than the defendant. Half of their participants also read a description of fingerprints and the methods used by fingerprint examiners in comparing a latent print to a suspect’s print. Finally, the expert testified about potential error rates but either discounted them, saying he himself had never made a mistake and that the chances that the defendant had not left the fingerprint were “infinitesimal,” or acknowledged them, saying that research has shown that fingerprint examiners make mistakes and that there is some possibility that the fingerprint was not left by the defendant. Only the presence of error information influenced likelihood of guilt ratings; when the expert discounted the possibility of error, participants rated it as more likely that the defendant was guilty and with greater confidence than when the expert acknowledged the possibility of error. Ratings of the probability that the defendant was the source of the evidence were also higher when the expert discounted than when he acknowledged the possibility of error. The form of the match testimony influenced only confidence in likelihood of guilt ratings; participants had greater confidence
when they had read the bolstered individualization testimony than the simple individualization testimony.

It is possible that participants were not more broadly influenced by differences in how the match was communicated because the manner in which they perceived the information was different than the manner in which it was intended. McQuiston-Surrett and Saks (2008) compared the language the American Board of Forensic Odontology (ABFO) recommended that bitemark experts use in their testimony to express varying levels of certainty to the manner in which undergraduates interpreted the language. According to the official ABFO guidelines, conclusions made to a “reasonable scientific certainty” were the strongest conclusions, indicating that there was “no reasonable probability of error,” followed by “probable,” “consistent with,” and “matching.” According to the ABFO, a match meant only that there was some correspondence between the source and the crime scene evidence but was not a statement of individualization. However, when undergraduates were asked to estimate the certainty that each statement represented about the likelihood that the evidence came from the suspect, they estimated that a match represented the greatest certainty (86% certain), followed by “consistent with,” “reasonable scientific certainty,” and “probable” (which they rated as representing a conclusion made with only 57.4% certainty).

**Diagnosticy**

Expert testimony may include information about the diagnosticity of a match (the likelihood that a source has been identified, given that a match has been found); however, jurors may not appropriately consider diagnosticity. Kaasa et al. (2007) investigated the extent to which jurors consider the diagnosticity of a forensic identification—here, bullet lead analysis—in evaluating the defendant’s guilt and the strength of the prosecution’s case. They presented mock
jurers with a homicide case including expert testimony regarding a bullet lead match that was presented as diagnostic (matched the defendant sample and not a control sample), non-diagnostic (just as likely to match the control sample as the defendant sample), or of unknown diagnosticity (no additional information). They found that participants did not consistently account for diagnosticity: participants rated the prosecution’s case as equally strong in all three conditions, and all three were rated as stronger than the control condition, where no bullet lead match was presented. Ratings of the probability that the defendant was guilty were higher when the match was diagnostic than in any other condition, suggesting some consideration of diagnosticity. Participants still overvalued a non-diagnostic match, however: even when the bullet was equally likely to match the control sample as the defendant’s sample, participants still thought it was more likely that the defendant was guilty than when no bullet lead match was presented.

Kaasa et al. (2007) also had participants deliberate after giving their initial judgments. They found that deliberation did aid participants in evaluating diagnosticity more appropriately: participants’ ratings of both the strength of the case and the probability that the defendant was guilty were higher when the match was diagnostic than in any other condition and the non-diagnostic match was no longer significantly different than control. Further analysis indicated that the results were largely driven by participants who reported that they were confident in their statistical abilities: confident participants’ judgments reflected that they considered whether the match was diagnostic, while unconfident participants’ judgments were similar across all experimental conditions. These results suggest that jurors may be able to differentiate between diagnostic and non-diagnostic information on a basic level, but their ability to do so may be determined, at least in part, by their comfort with mathematics and statistics and by their ability to discuss the evidence with other jurors.
Individual differences: Need for Cognition. Kaasa et al. (2007) found differences in evidence evaluation based on comfort with math and science, while Garrett and Mitchell (2013) found some differences based on numeracy; however there may be broader individual differences in the extent to which jurors are able and/or motivated to evaluate the validity of forensic evidence and to weigh it appropriately. One particular difference—which may relate to numeracy and general comfort with mathematics—is the Need for Cognition. Need for Cognition (NC) refers to the extent to which an individual tends to engage—and to enjoy engaging—in the effortful and analytical processing of information (Cacioppo & Petty, 1982; Cohen, Stotard, & Wolfe, 1955). Individuals high in NC are more likely to expend mental energy on the systematic processing of information, while those low in NC are more likely to use less effortful, shallower forms of processing and show a greater reliance on heuristics (Cacioppo, Petty, & Morris, 1983). NC affects judgments and decision-making across a wide range of domains (Cacioppo, Petty, Feinstein, & Jarvis, 1996), including the evaluation of persuasive messages (e.g., Cacioppo et al., 1983), development of expectations for personal performance (Dickhauser & Reinhard, 2006), response to commercial advertisements (Haugtvedt, Petty, & Cacioppo, 1992), organization of ongoing behavior into actions (Lassiter, Briggs, & Bowman, 1991), and use of the media in gathering information (Cacioppo et al., 1996).

In the legal arena, some evidence suggests that NC affects what types of evidence most influence jurors (e.g., Bornstein, 2004), as well as the manner and extent to which jurors systematically evaluate the evidence (e.g., DeWitt, Richardson, & Warner, 1997; Kassin, Reddy, & Tulloch, 1990; Leippe, Eisenstadt, Rauch, & Seib, 2004; McAuliff & Kovera, 2008). While certain types of evidence may generally be more persuasive than others (see above; Lieberman et al., 2008; Skolnick & Shaw, 2001), persuasiveness may also vary as a result of NC. Perhaps
related to their reduced likelihood of systematically evaluating evidence, people low in NC may be more influenced by anecdotal evidence than experimental or statistical evidence. Bornstein (2004) found that low NC mock jurors were more persuaded by a defense expert when he presented anecdotal evidence than when he did not. They also rated the defendant as less likely to have caused the plaintiff’s injuries. High NC jurors were not affected by the presence of anecdotal evidence.

NC may also affect jurors’ ability to appropriately evaluate and weigh scientific evidence. Jurors may be less likely to carefully consider the content of expert testimony or to scrutinize attorney arguments when they are low in NC than when they are high in NC (DeWitt et al., 1997; Leippe et al., 2004). In addition, low NC jurors may be more likely to rely on heuristics (e.g., “experts are trustworthy”) when determining the strength of evidence. Thus, high NC jurors may be more likely to take scientific validity into account when evaluating scientific evidence and may be more persuaded by more valid types of forensic evidence; low NC jurors may not make this distinction, relying only on the fact that an expert is present. For example, McAuliff and Kovera (2008) examined whether NC had an impact on the sensitivity of mock jurors to methodological flaws in the scientific research presented by an expert in a hostile work environment case. They presented high NC and low NC mock jurors with expert testimony wherein the plaintiff’s expert described a study that was either internally valid or was missing a control group. They also manipulated ecological validity (whether the study’s participants were undergraduates or were similar to the employees at the plaintiff’s workplace) and the general acceptance of the study (whether it was published in a peer-reviewed journal). High NC mock jurors were more likely to find for the plaintiff and rated the expert evidence as higher in quality when the study presented was internally valid that when it lacked a control group. By contrast,
low NC mock jurors were not influenced by the manipulation of internal validity. However, neither high nor low NC mock jurors were affected by the manipulation of the ecological validity or the general acceptance of the study, suggesting that even high NC participants were not sensitive to all variations in evidence quality.

Despite being likely to expend more effort in evaluating testimony, it is possible that such efforts may actually lead high NC jurors to be more likely to come to conclusions that are in line with what they already believe. Evidence for this possibility comes from studies showing that high NC people are more influenced by information received earlier than information received later (Haugtvedt & Petty, 1992; Kassin et al., 1990). For example, the order in which evidence and arguments are presented may influence jurors high and low in NC differently, particularly when evidence is ambiguous: low NC jurors are more influenced by attorney arguments that follow the presentation of evidence, while high NC jurors are more influenced by arguments given prior to the presentation of evidence (Kassin et al., 1990). Additionally, high NC individuals are more influenced by primacy effects (Haugtvedt & Petty, 1992), perhaps in part because they are more likely to consider a particular argument (e.g., an attorney argument) while evaluating the evidence. Unfortunately, their greater predisposition to consider prior evidence or preconceived notions may actually make people high in NC more susceptible to biased assimilation effects (Haugtvedt & Petty, 1992; Kassin et al., 1990), making them more likely to evaluate subsequent information in light of the information they already have. Thus, it is possible that high NC jurors may be more likely than low NC jurors to evaluate the evidence in line with their pretrial beliefs.

High NC jurors are better able to comply with judicial instructions, but may also be more likely to overcorrect for potential sources of bias the instructions may alert them to. Sommers
and Kassin (2001) presented high and low NC jurors with a trial summary in which wiretapping evidence was either not mentioned, was admitted and found admissible, or was admitted and then ruled inadmissible either on the basis of due-process violations or on the basis of unreliability. For low NC jurors, the differences in verdict among conditions did not reach significance; for high NC jurors, convictions were higher when the wiretapping evidence was admitted and found admissible than in any other condition. Although overcorrection was not found in verdict choices (i.e., verdict choices did not differ in the admissible and unreliable inadmissible conditions), it was found in high NC jurors’ ratings of the probability that the defendant committed the crime. When the wiretapping evidence had been ruled inadmissible due to unreliability, high NC jurors’ ratings of the probability that the defendant committed the crime were significantly lower than when no wiretapping evidence was mentioned. No evidence for overcorrection was found for low NC jurors. The authors suggest that high NC jurors, in part due to their greater motivation to reach a just verdict, made an active attempt to engage in bias correction. In essence, they assumed that the evidence had biased them against the defendant and attempted to correct for the bias by downward-adjusting their probability estimates. These results suggest that when jurors are alerted to the potential unreliability of evidence, high NC jurors may engage in bias correction and may overcorrect, resulting in lower perceptions of guilt than if they had not been presented with the evidence at all. Thus, for high NC jurors, cross-examination and other legal safeguards that focus on the unreliability of forensic science may result in lower perceptions of guilt than if no forensic evidence was presented at all.

While cross-examination may cause high NC jurors to overcorrect for potential bias, it may be helpful for low NC jurors by prompting them to consider the weaknesses of the evidence, including potential problems with reliability and validity. Low NC people have a reduced
tendency to spontaneously engage in deeper processing; thus, cross-examination may be useful in encouraging low NC jurors to carefully consider the evidence, whereas high NC jurors may do so without prompting (Salerno & McCauley, 2009).
CHAPTER 5: THE EFFECTIVENESS OF CROSS-EXAMINATION

Even if cross-examination or other traditional legal safeguards may influence high or low NC jurors’ perceptions of guilt in some circumstances, it is not clear whether they are generally effective in influencing jurors’ verdicts or their perceptions of particular pieces of evidence (e.g., Kovera et al., 1999; Lieberman et al., 2008; McQuiston-Surrett & Saks, 2009; Salerno & McCauley, 2009). In fact, some evidence suggests that traditional legal safeguards—such as cross-examination, opposing expert witnesses, and judicial instruction—may not effectively educate jurors about invalid science and may not help jurors to more accurately evaluate scientific evidence. Kovera et al. (1999) examined the influence of two different types of cross-examination of an expert in a hostile work environment case. The expert testified regarding the results of a study she conducted on sexual harassment. The researchers varied the quality of the study by manipulating general acceptance, ecological validity, and construct validity. Scientifically naïve cross-examination asked for more information about the effects of the study and about the application of the study to the plaintiff’s case. Scientifically informed cross-examination additionally questioned the methodology of the study, including targeted questions about ecological and construct validity. Although the authors hypothesized that the scientifically informed cross-examination would sensitize jurors to differences in the quality of the study, they found no effect of cross-examination type on perceptions of the defendant’s liability or the expert’s credibility.

 Interestingly, prior research examining the influence of cross-examination of forensic identification experts has found that while cross-examination does not directly affect verdicts, it may influence related judgments of the evidence (Koehler, 2011; Lieberman et al., 2008; McQuiston-Surrett & Saks, 2009). Koehler (2011) examined whether the inclusion of cross-
examination of a shoeprint expert influenced perceptions of guilt, the strength of the evidence, and the likelihood that the defendant was the source of shoeprints left at the scene of a burglary. He found no effect of cross-examination on verdicts or ratings of the probability that the defendant committed the burglary or was the source of the shoeprint. However, cross-examination influenced ratings of the strength of the evidence: mock jurors who read cross-examination focused on the potential for error in shoeprint analysis rated the evidence as weaker and as less convincing than participants who read only the direct testimony of the expert.

McQuiston-Surrett and Saks (2009) examined the effects of judicial instruction and cross-examination of the testimony of an expert in microscopic hair analysis. They found that neither safeguard had a significant impact on verdicts or jurors’ estimates of the probability that the hair belonged to the defendant; however, both increased jurors’ self-ratings of how well they understood the science presented. McQuiston-Surrett and Saks’s results suggest that cross-examination may make jurors feel that they have become better science consumers but without affecting their perceptions of the probative value of the evidence in a particular case.

The influence of cross-examination may, to some extent, be determined by the primary target of the cross-examination. In a study examining DNA evidence only, Lieberman et al. (2008, Study 3) compared the effectiveness of cross-examination targeting the expert to cross-examination targeting the quality of the evidence itself. In the expert-focused cross-examination, the defense attorney questioned the expert’s credentials and experience, and whether the money he received in exchange for his testimony might have influenced his findings. In the evidence-focused cross-examination, the defense attorney expressed concern about the potential for evidence contamination, the potential for error in judgment, and the existence of wrongful convictions that involved DNA evidence. Lieberman et al. also manipulated the reliability of the
lab in which the expert worked. In the reliable-lab condition, the lab was not affiliated with the police department, was accredited, conducted regular proficiency testing, used control samples, and had trained the expert in scientific procedures. In the unreliable-lab condition, the lab was affiliated with the police department, the expert had been trained as a police officer (not a scientist), the lab was not accredited, and proficiency testing and control samples were not mentioned. Lieberman et al. found no main effect of type of cross-examination on participants’ perceptions of guilt; however, they found that expert-focused cross-examination was more effective when the expert’s lab was presented as reliable, while evidence-focused cross-examination was more effective when the lab was presented as unreliable. Their results suggest that the effectiveness of the different cross-examination types may, at least to some degree, be affected by reliability. If so, then it is possible that different cross-examination types may be more effective with more versus less valid forms of evidence. Thus, expert-focused cross-examination may be more effective in challenging valid forensic science (e.g., DNA evidence), while evidence-focused cross-examination may be more effective in challenging less valid forensic science (e.g., bitemark evidence).
CHAPTER 6: STUDY ONE OVERVIEW

In Study 1, I examined mock jurors’ beliefs about science in general and about the reliability of different types of forensic evidence in specific. I also examined whether they were able to appropriately weigh different types of forensic science presented as evidence in a criminal trial. I was interested in whether their beliefs influenced their perceptions and judgments, and whether these beliefs, if misinformed, could be countered by cross-examination. Further, I examined whether NC influenced participants’ judgments of the different types of evidence and the effectiveness of different types of cross-examination.

Briefly, after rating the reliability of a number of different types of evidence, participants read a trial transcript including either no expert (control group) or the direct and cross-examination of a forensic expert. I used four different types of forensic science evidence, selected to vary in validity based on the findings of the NAS Report: DNA, fingerprint, toolmark, and bitemark evidence. Because prior research has found differences in the manner in which forensic identifications and non-identifications influence juror decision-making (Garrett & Mitchell, 2013; Lieberman et al., 2008; McAllister & Bregman, 1986), the evidence was presented as either matching or not matching the defendant. I also examined whether cross-examination focused on the evidence or the expert was more successful at countering naïve beliefs and inducing jurors to view the less valid types of evidence more critically. Considering that most opposing attorneys providing effective assistance of counsel would offer at least a brief cross-examination, I compared these two types of cross-examination to a weak, non-substantive cross-examination rather than no cross-examination.
CHAPTER 7: STUDY ONE METHOD

Participants

Participants were 853 adults recruited online via Amazon’s Mechanical Turk (MTurk). Participants recruited through MTurk complete studies or other tasks online in exchange for a small amount of compensation that they can either apply towards Amazon Gift Certificates or transfer directly to their bank account. MTurk automatically screens out participants under the age of 18, and has mechanisms in place to ensure that participants do not complete the same study more than once. Based on prior research, we know that participants recruited and run online through MTurk are comparable (and sometimes preferable) to participants completing studies in person; indeed, the demographic characteristics of the MTurk subject pool and the quality of MTurk participants suggests that the sample is diverse and that participants provide high-quality data (Buhrmester, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013; Paolacci, Chandler, & Ipeirotis, 2010; Sprouse, 2011). I restricted availability of the study to participants located in the United States who had received a high rate of approval from other researchers.

Of the participants completing Study 1, 159 failed manipulation and/or attention checks (see Materials) and were excluded, leaving a final sample of 694 participants. The age of participants ranged from 18 to 82 years ($M = 34.39$, $SD = 11.84$). There were 247 male participants (35.6%) and 443 female participants (63.8%); 4 did not provide a gender. Most were White (82%), but the sample also included Hispanic (4.6%), Black (5.3%), and Asian participants (4.5%). The vast majority of participants (89.6%) had completed at least some college; almost a third (31.6%) had a Bachelor’s degree and 15.6% had completed at least some graduate school. All but twelve participants were U.S. citizens (excluding these participants did
not change the pattern of results, so I retained them). Participants were compensated $0.80 for their participation; an amount comparable to that used by other researchers on MTurk for similar periods of time.

**Design**

The design of the study was a 4 (forensic evidence type: DNA, fingerprint, bitemark, toolmark) x 2 (evidence match type: match, non-match) x 3 (cross-examination type: evidence-focused, expert-focused, non-substantive) + 1 (no forensic evidence control group) between-subjects factorial design.

**Materials**

**Attitudes towards science.** To assess general attitudes towards science, I used the questions used by the National Science Board (2004, 2006; J. D. Miller, 2004) in their surveys regarding belief in the benefits of science and reservations about science (see Table 2). Four questions assessed belief in the benefits of science and three assessed reservations about science. All items were rated on a scale of 1 (strongly disagree) to 4 (strongly agree). The first four items were summed to create a “beliefs-in-science” composite variable and the second three were summed to create a “reservations-about-science” composite variable; thus, scores for *beliefs-in-science* ranged from 4-16, while scores for *reservations-about-science* ranged from 3-12.

**Pretrial naïve beliefs about forensic science.** Questions regarding naïve beliefs about evidence types were based on those used by Hans et al. (2011), who asked participants to rate the reliability of different evidence types on a scale from 1 (not at all reliable) to 5 (extremely reliable). However, Hans et al. used only one type of forensic science (DNA) and compared it to four types of non-scientific evidence (police, expert, victim, and eyewitness). I added the three other types of forensic science to be examined in Study 1 (fingerprint, toolmark, and bitemark). I
also added two additional types of evidence (polygraph and confession) so that it was less obvious that I was interested in forensic science specifically. The order in which types of evidence were listed on the study site was randomized. Participants were also given the option of saying that they had never heard of the type of evidence in lieu of giving a reliability rating. I used the pretrial reliability rating for the specific type of evidence experimental participants read about in their assigned transcript in analyses involving specific naïve beliefs about forensic evidence.

**Transcripts.** For the purposes of the study, I created trial transcripts based on a fictitious homicide case; expert testimony was adapted from several trial transcripts involving the four types of forensic evidence and then equated as much as possible. The transcripts included judicial instructions, opening and closing statements of the prosecutor and defense attorney, testimony from the investigating detective and testimony from a forensic science expert in the experimental conditions or testimony from a witness putting the defendant and victim together in control conditions (note that the same information contained in the witness’s testimony was also included in the detective’s testimony so that the control group did not receive any additional information). A number of steps were taken to equalize the transcripts across conditions. For example, opening statements and the testimony of the investigating officer was the same in all conditions. The physical evidence on which testing was conducted was also the same in all conditions: a length of rubber tubing that was used to strangle the victim. In DNA conditions, the expert tested saliva from a bitemark on the tubing; in fingerprint conditions, he tested a fingerprint on the tubing; in toolmark conditions he tested whether a knife owned by the defendant was used to cut the tubing (the knife was found in all conditions, but only analyzed in this condition); and in bitemark conditions he tested whether a bitemark on the tubing matched
the defendant. Transcripts differed only in whether the forensic science expert testified, which type of evidence he testified about and his ultimate conclusion, and which side presented the testimony (he was a prosecution expert when the evidence matched and a defense expert when it did not). The expert’s testimony was referred to by attorneys only in closing arguments (opening statements were brief and did not mention specific pieces of evidence), and then only briefly.

Within the expert testimony, I attempted to equalize all elements not directly related to the type of science analyzed. Forensic expert direct testimony contained the following elements: 1) expert credentials, which contained the same questions in all conditions with answers differing only when responses were of necessity linked to the particular type of science (e.g., “I have taken numerous training courses in latent print analysis” vs. “I have taken numerous training courses in forensic odontology”); 2) basic overview of the field and how identifications are made; and 3) specifics of the identification in the instant case. All experts made their conclusions regarding source “to a reasonable degree of scientific certainty.” Length was also kept largely the same across evidence conditions.

The non-substantive cross-examination questioned the certainty of the expert (whether the expert was 100% sure) and whether the fact that the expert’s opinion that the defendant was (or was not) the source of the evidence meant that he necessarily did (or did not) commit the crime. The content of the non-substantive cross-examination was also included in the expert- and evidence-focused cross-examination. The expert-focused and evidence-focused cross-examination conditions were adapted from the materials used by Lieberman et al. (2008, Study 3). Lieberman et al. used only DNA evidence; thus the materials were adapted for use in the other evidence conditions. Expert-focused cross-examination centered on the expert’s credentials and the possibility that the expert was a hired gun (either on the side of the prosecution or the
defense). The expert-focused cross-examination was kept largely the same across conditions with the exception of the names of the credentials attacked. Evidence-focused cross-examination centered on the presence or absence of validation studies supporting the technique, problems with the technique itself, the potential for bias in the expert’s conclusions, and the possibility of contamination. The evidence-focused cross-examination was equalized to the greatest extent possible; however, questions attacking the methods used in the analysis of a particular type of evidence and responses differed by necessity.

**Need for Cognition Scale.** I used the short version of the Need for Cognition Scale to assess NC (Cacioppo, Petty, & Kao, 1984). The scale contains 18 items, half of which are reverse coded. Items are scored on a 5-point Likert scale where 1 = extremely uncharacteristic; 2 = somewhat uncharacteristic; 3 = uncertain; 4 = somewhat characteristic; 5 = extremely characteristic; thus, once items are reversed, total possible scores range from 18-90. Once NC scores were calculated, I divided the scores into thirds based on percentiles to create a low NC, moderate NC, and high NC group. Mean scores were 54.83 (SD = 10.30) for the low NC group, 70.46 (SD = 2.27) for the moderate NC group, and 81.00 (SD = 4.12) for the high NC group.

**Manipulation checks.** Six multiple choice questions assessed attention to the materials and to the experimental manipulations (see Appendix B). The first three assessed general attention to the trial materials. The last three assessed attention to the experimental manipulations, the first asking whether an expert testified, the second asking what type of expert, and the third asking whether the expert linked the defendant to the evidence. I also asked participants to give open-ended descriptions of the most important parts of the trial and of the expert’s testimony in order to catch any problems with the questions. Indeed, upon inspection of the descriptions, I discovered that the final manipulation check was actually interpreted by some
participants as a perception of guilt or evidence strength measure, rather than asking simply whether the expert found that the evidence matched. For example, both of these participants said that the expert did not link the defendant: “The most important part of his testimony were linking the knife found in the defendants [sic] car to the cuts made in the rubber tubing. But I feel like the most damaging part of his testimony (for the prosecution) was when they were talking about the cuts not being ‘exact’ and insinuating that personal opinion may have influenced the experts [sic] findings” and “There is not concrete evidence backing the technique used to analyze the tool marks. The marks were not confirmed to be a 100% match to the tool.” Therefore, I did not exclude participants based on their responses to this manipulation check question and it was not used in Studies 2 and 3.

**Dependent Measures**

Regarding perceptions of guilt, I asked participants to render dichotomous verdicts (guilty, not guilty) and to rate their confidence in those verdicts on a scale of 1 (not at all) to 7 (very). I also asked participants to rate the likelihood that the defendant murdered the victim and—in experimental conditions—the likelihood that the defendant (or his knife) was the source of the evidence, both as a percentage from 1 (definitely did not/is not) to 100 (definitely did/is).

Regarding perceptions of the expert, I asked participants to rate where the expert fell on a series of adjective scales (e.g., “competent-incompetent”). I averaged scores on these items to form an “expert credibility” composite variable. Regarding perceptions of the evidence, I asked participants a series of questions about the reliability and validity of the evidence they viewed (e.g., “The forensic identification technique used by expert witness Williams was based on good scientific principles”), as well as the appropriateness of the evidence to the case and how helpful the evidence was, each on a scale of 1 (disagree strongly) to 6 (agree strongly). Three of these
items were reverse coded (see Appendix B). I averaged scores on these items to form “evidence validity” and “evidence helpfulness” composite variables. Finally, I assessed participants’ perceived understanding of the expert’s testimony and his forensic technique, each on a scale of 1 (not at all well) to 7 (very well).

**Procedure**

Participants signed up on the MTurk site and then were directed to an external study website (www.psychsurveys.org) that contained all materials and questionnaires. The study site randomly assigned participants to condition. After completing informed consent (see Appendix A), participants answered the preliminary questions assessing attitudes towards science and beliefs about evidence reliability. They then read their assigned transcript and responded to the manipulation check questions, and then the post-trial perceptions of guilt and perceptions of the evidence and expert questions. Once they had completed the post-trial questions, they completed the Need for Cognition Scale and provided basic demographic information. When they had completed the study, they were given a unique ID number which they could enter into the MTurk site and which I used to check whether they had completed the study. Once I was able to verify that their ID number was valid, I released their payment.

**Hypotheses**

H1: DNA evidence will be viewed as most reliable, followed by fingerprint evidence, toolmark evidence, and bitemark evidence.

H2: Guilty verdicts and ratings of the likelihood that the defendant murdered the victim will be higher in the match conditions than in the control condition, but the non-match conditions will not differ from the control condition.
H3: There will be an interaction between the type of evidence and whether the evidence matches the defendant on perceptions of guilt such that participants will differentiate between evidence types when the evidence matches but there will be no differences between evidence types when it does not.

H4: Regardless of whether the evidence is presented as matching the defendant, the expert will be viewed as most credible and the evidence as highest quality in DNA conditions followed by fingerprint, toolmark, and bitemark conditions, and in non-substantive than substantive cross-examination conditions.

H5: Prior beliefs about the reliability of the relevant type of evidence will significantly predict perceptions of guilt and perceptions of the expert and evidence.

H6: Attitudes about science will not significantly predict perceptions of guilt, but will predict perceptions of the expert and the evidence.

H7: Pretrial beliefs will function as better predictors of perceptions of guilt and perceptions of the expert and evidence for high NC than low NC participants.

H8: Low NC participants will be equally influenced by both types of substantive cross-examination (expert-focused and evidence-focused), while high NC participants will differentiate between types of cross-examination such that expert-focused cross-examination will be more effective with more valid types of evidence, while evidence-focused cross-examination will be more effective with less valid types of evidence.
CHAPTER 8: STUDY ONE RESULTS

Due to the large amount of data resulting from Study 1, I organized my results based on my hypotheses. For each hypothesis, I summarize the findings supporting or refuting the hypothesis before presenting the analyses. I additionally put the vast majority of means and standard deviations in tables in order to increase the readability of the results. In my analyses, I used logistic regression for verdicts and ANOVA or MANOVA for all other analyses with Tukey tests for all post hoc follow-ups. A disproportionate number of participants reported that they had never heard of toolmark or bitemark evidence (see Table Y); thus, in order to have more even Ns when possible, I only included pretrial beliefs in my analysis of hypotheses related to pretrial beliefs. With the exception of my comparisons to the control group, all models included evidence type, match type, and cross-examination type unless otherwise specified. In my analyses of pretrial beliefs, I examined ratings of the specific types of evidence participants read about (specific naïve beliefs), and participants’ more general positive and negative attitudes about science (beliefs-in-science, and reservations-about-science, respectively). For these analyses, I used ANCOVA or MANCOVA to analyze continuous dependent measures with beliefs included as a covariate.

Hypothesis 1: DNA evidence will be viewed as most reliable, followed by fingerprint evidence, toolmark evidence, and bitemark evidence.

My first hypothesis was supported except that bitemark evidence was rated as more reliable than toolmark evidence. Mean reliability ratings were highest for DNA, followed by fingerprint, bitemark, and toolmark evidence; all significantly differed from each other at the $p < .001$ level (see Table 3). However, even the lowest rated forensic science was given significantly
higher ratings than any other type of evidence. Polygraph evidence was rated as the least reliable, followed by eyewitness evidence.

In examining pretrial beliefs, I also examined participants’ general beliefs about science. More than 85% of participants agreed or strongly agreed with each of the individual items that comprised the beliefs-in-science composite, while fewer than 25% of participants agreed or strongly agreed with each of the items that comprised the reservations-about-science composite (see Table 2). In fact, the mean beliefs-in-science score was 13.04 ($SD = 1.93$), while the mean reservations-about-science score was 5.67 ($SD = 1.83$). Further, there was a strong negative association between beliefs-in-science and reservations-about-science, $r(688) = -0.40$, $p < .001$.

**Hypothesis 2:** Guilty verdicts and ratings of the likelihood that the defendant murdered the victim will be higher in match conditions than in the control condition, but the non-match conditions will not differ from the control condition.

My second hypothesis was supported in that a match had a greater influence than a non-match, and verdict choice in non-match conditions did not differ from the control group; however, a non-match did reduce estimates of the probability that the defendant murdered the victim.

The percentage of guilty verdicts was significantly higher in match conditions (62.9%) than it was in the control group (17.6%), $\chi^2(1) = 36.62$, $p < .001$, $\Phi = .32$ (95% CI: .21, .39); however, the percentage of guilty verdicts in non-match conditions (9.4%) did not differ significantly from the control group, $\chi^2(1) = 3.20$, $p = .07$, $\Phi = -.09$ (95% CI: -.24, .02). When the evidence matched the defendant, guilty verdicts were significantly more likely for participants viewing each of the four types of evidence compared to the control group (all $ps < .001$; for
percentages in experimental groups see Table 4). An ANOVA revealed that estimates of the probability that the defendant murdered the victim were significantly influenced by whether the participant was in the control versus the match or non-match conditions, $F(2, 690) = 147.06, p < .001, \eta^2_p = .30$. Post hoc Tukey tests indicated that probability estimates were significantly higher in match conditions ($M = 76.36, SD = 22.84$) than in the control condition ($M = 54.55, SD = 24.98; p < .001, M_{diff} = 21.81$ [95% CI: 13.30, 30.33]). They were significantly lower in non-match conditions ($M = 43.96; SD = 24.95$) than in the control condition ($p = .01, M_{diff} = -10.59$ [95% CI: -19.08, -2.10]), although the difference was considerably smaller.

**Hypothesis 3:** There will be an interaction between the type of evidence and whether the evidence matches the defendant on perceptions of guilt such that participants will differentiate between evidence types when the evidence matches but there will be no differences between evidence types when it does not.

My third hypothesis was supported for verdicts and estimates of the probability that the defendant murdered the victim. For estimates of the probability that the defendant was the source of the evidence, however, there was some differentiation between evidence types even when the evidence did not match. Further, I found only main effects for verdict confidence.

**Verdicts.** I first tested a logistic regression model including evidence type, match type, and cross-examination type and all two-way interactions in order to examine the differences between my experimental groups in verdicts. The model was significant, $\chi^2(17) = 247.94, p < .001, R^2_{nag} = .44$. I found a significant main effect of match type, Wald $\chi^2(1) = 17.58, p < .001$, $OR = 0.08$ (95% CI = 0.03, 0.27), in addition to a significant interaction between evidence type and match type, Wald $\chi^2(3) = 12.85, p = .005$. There was not a significant main effect of
evidence type, Wald $\chi^2(3) = 6.34$, $p = .10$, or of cross-examination type, Wald $\chi^2(2) = 0.74$, $p = .69$, and no other interactions reached significance (all $ps > .3$). Guilty verdicts were significantly more likely in match conditions than in non-match conditions, $\chi^2(1) = 201.25$, $p < .001$, $\Phi = .56$. There was also a significant effect of evidence type on verdict in match conditions, $\chi^2(1) = 18.93$, $p < .001$, $\Phi = .25$, while verdicts did not significantly differ in non-match conditions, $\chi^2(1) = 2.51$, $p = .47$, $\Phi = .09$ (see Table 4 for the full verdict breakdown by evidence and match type).

**Verdict confidence.** Both the type of evidence participants read about and whether the evidence matched the defendant influenced verdict confidence: an ANOVA revealed significant main effects of evidence type, $F(1, 619) = 4.96$, $p = .002$, $\eta_p^2 = .02$, and of match type, $F(1, 619) = 10.62$, $p = .001$, $\eta_p^2 = .02$. There was no effect of cross-examination type, $F(1, 619) = 0.31$, $p = .73$, $\eta_p^2 = .001$, and no interactions reached significance (all $ps > .5$). Follow up analyses revealed that verdict confidence was significantly higher in DNA conditions than it was in toolmark ($p = .005$, $M_{\text{diff}} = 0.64$ [95% CI: 0.15, 1.14]) or bitemark conditions ($p = .02$, $M_{\text{diff}} = 0.56$ [95% CI: 0.08, 1.05]). Verdict confidence tended to be higher in fingerprint conditions than it was in toolmark conditions, but the difference did not reach significance ($p = .08$, $M_{\text{diff}} = 0.46$ [95% CI: -0.03, 0.96]). Participants were also more confident in their verdicts when the evidence matched than when it did not match the defendant (see Table 5).

**Murder probability estimates.** For participants’ estimates of the probability that the defendant murdered the victim, an ANOVA revealed a significant main effect of match type, $F(1, 618) = 292.28$, $p < .001$, $\eta_p^2 = .32$, and a significant interaction between evidence type and match type, $F(3, 618) = 4.44$, $p = .004$, $\eta_p^2 = .02$. There was not a significant main effect of evidence type, $F(3, 618) = 1.15$, $p = .33$, $\eta_p^2 = .006$, or cross-examination type, $F(2, 618) = 0.39$, $p = .68$, $\eta_p^2 = .001$, and no other interactions reached significance (all $ps > .4$). Estimates were
higher when the expert testified that the evidence matched than when he testified that it did not (see Table 6). Follow-up tests indicated that there was a significant difference between evidence types in match conditions, $F(3, 309) = 5.24, p = .002, \eta^2_p = .05$, but no significant difference between evidence types in non-match conditions, $F(3, 325) = 0.96, p = .41, \eta^2_p = .01$. Within match conditions, post-hoc Tukey tests indicated that estimates were significantly higher in the DNA condition than they were in the toolmark ($p = .01, M_{diff} = 10.99 [95\% CI: 1.70, 20.29]$) or bitemark conditions ($p = .002, M_{diff} = 13.09 [95\% CI: 3.91, 22.26]$). The fingerprint condition did not differ significantly from any other condition (all $ps > .6$).

**Source probability estimates.** For estimates of the probability that the defendant was the source of the evidence, an ANOVA revealed a significant main effect of match type, $F(1, 617) = 2021.89, p < .001, \eta^2_p = .77$, a significant interaction between evidence type and match type, $F(3, 617) = 15.51, p < .001, \eta^2_p = .07$, and a significant interaction between match type and cross-examination type, $F(2, 617) = 16.19, p < .001, \eta^2_p = .05$. There was not a significant main effect of evidence type, $F(3, 617) = 1.09, p = .35, \eta^2_p = .005$, or cross-examination type, $F(2, 617) = 1.43, p = .24, \eta^2_p = .005$, and no other interactions reached significance (all $ps > .6$). Estimates were higher when the expert testified that the evidence matched than when he testified that it did not (see Table 6). Follow-up analysis indicated that in non-match conditions, estimates of the likelihood the defendant was the source of the evidence were significantly higher in the toolmark condition than they were in the DNA condition ($p = .006; M_{diff} = 11.41 [95\% CI: 2.66, 20.16]$). They were significantly higher in the evidence-focused cross-examination conditions than they were in the non-substantive cross-examination ($p < .001, M_{diff} = 13.19 [95\% CI: 6.34, 20.04]$) or the expert-focused cross-examination conditions ($p = .01, M_{diff} = 8.44 [95\% CI: 1.66, 15.22]$). In match conditions, estimates of the likelihood that the defendant was the source of the evidence
were significantly higher in the DNA condition than they were in toolmark \((p < .001; M_{\text{diff}} = 15.98 [95\% \text{ CI: } 9.28, 22.68])\) or bitemark conditions \((p < .001; M_{\text{diff}} = 10.87 [95\% \text{ CI: } 4.26, 17.48])\). They were also significantly higher in the fingerprint condition than they were in toolmark \((p < .001; M_{\text{diff}} = 12.90 [95\% \text{ CI: } 6.18, 19.62])\) or bitemark conditions \((p = .02; M_{\text{diff}} = 7.78 [95\% \text{ CI: } 1.15, 14.41])\). They were significantly lower in the evidence-focused cross-examination condition than they were in the non-substantive cross-examination condition \((p < .001; M_{\text{diff}} = -7.14 [95\% \text{ CI: } -12.42, -1.85])\), but were not significantly different from the expert-focused condition \((p = .09; M_{\text{diff}} = -4.70 [95\% \text{ CI: } -9.98, 0.58])\).

**Hypothesis 4:** Regardless of whether the evidence is presented as matching the defendant, the expert will be viewed as most credible and the evidence as highest quality in DNA conditions followed by fingerprint, toolmark, and bitemark conditions, and in non-substantive than substantive cross-examination conditions.

My hypothesis regarding perceptions of the expert and evidence was only partially supported, and I found some unexpected results. DNA evidence was viewed as most valid, but the DNA and fingerprint experts were viewed as equally credible and both as more credible and valid than the toolmark or bitemark experts (who did not differ). Additionally, both types of substantive cross-examination reduced perceptions of expert credibility, while only evidence-focused cross-examination reduced perceptions of evidence validity. Unexpectedly, the presence of a match also influenced perceptions of the expert and the evidence: when the expert testified that the evidence matched, participants rated the expert as more credible and the evidence as higher quality.
**Perceptions of expert credibility.** An ANOVA revealed significant main effects of evidence type, $F(3, 614) = 8.30, p < .001, \eta^2_p = .04$, match type, $F(1, 614) = 8.03, p = .005, \eta^2_p = .01$, and cross-examination type, $F(2, 614) = 16.25, p < .001, \eta^2_p = .05$, on my composite measure of expert credibility; there were no significant interactions (all $ps > .2$). Ratings of expert credibility were significantly higher in the DNA condition than they were in the toolmark ($p < .001, M_{diff} = 0.42 [95\% CI: 0.16, 0.69]$) or the bitemark conditions ($p < .001, M_{diff} = 0.41 [95\% CI: 0.15, 0.67]$) (see Table 7 for means). They were also significantly higher in the fingerprint condition than they were in the toolmark condition ($p = .04, M_{diff} = 0.28 [95\% CI: 0.01, 0.54]$), and marginally higher than in the bitemark condition ($p = .05, M_{diff} = 0.26 [95\% CI: -0.0002, 0.52]$). The expert was also viewed as more credible when he testified that the evidence matched the defendant than when he testified that it did not. Regarding cross-examination the expert was viewed as more credible in the non-substantive cross-examination condition than he was in either the expert-focused ($p = .007, M_{diff} = 0.27 [95\% CI: 0.06, 0.48]$) or evidence-focused ($p < .001, M_{diff} = 0.50 [95\% CI: 0.30, 0.71]$) cross-examination conditions. The expert-focused and evidence-focused conditions also differed significantly ($p = .02, M_{diff} = 0.24 [95\% CI: 0.03, 0.44]$).

**Perceptions of evidence quality.** A MANOVA revealed significant main effects of evidence type, Wilks’ $\lambda = .91, F(6, 1226) = 10.29, p < .001, \eta^2_p = .05$, match type, Wilks’ $\lambda = .99, F(2, 612) = 3.58, p = .03, \eta^2_p = .01$, and cross-examination type, Wilks’ $\lambda = .91, F(4, 1224) = 14.25, p < .001, \eta^2_p = .04$, in addition to a significant interaction between match type and cross-examination type, Wilks’ $\lambda = .98, F(4, 1224) = 2.60, p = .04, \eta^2_p = .008$, on my composite measures of evidence validity and helpfulness. Univariate tests indicated that there was a significant main effect of evidence type on perceptions of both evidence validity, $F(3, 613) =$
20.45, $p < .001$, $\eta_p^2 = .09$, and evidence helpfulness, $F(3, 613) = 8.93, p < .001$, $\eta_p^2 = .04$. Post-hoc Tukey tests revealed that DNA evidence was viewed as more valid than fingerprint ($p = .01$, $M_{\text{diff}} = 0.27$ [95% CI: 0.04, 0.50]), toolmark ($p < .001$, $M_{\text{diff}} = 0.56$ [95% CI: 0.33, 0.79]), or bitemark evidence ($p < .001$, $M_{\text{diff}} = 0.61$ [95% CI: 0.39, 0.84]); fingerprint evidence was also viewed as more valid than toolmark ($p = .009$, $M_{\text{diff}} = 0.29$ [95% CI: 0.06, 0.52]) or bitemark evidence ($p = .001$, $M_{\text{diff}} = 0.34$ [95% CI: 0.11, 0.57]) (see Table 7 for means). DNA evidence was viewed as more helpful than toolmark ($p = .003$, $M_{\text{diff}} = 0.49$ [95% CI: 0.12, 0.85]) or bitemark evidence ($p < .001$, $M_{\text{diff}} = 0.65$ [95% CI: 0.29, 1.00]); fingerprint evidence was also viewed as more helpful than bitemark evidence ($p = .005$, $M_{\text{diff}} = 0.46$ [95% CI: 0.11, 0.82]).

There was a main effect of match type on perceptions of the validity of the evidence, $F(1, 613) = 7.07, p = .008$, $\eta_p^2 = .01$, but not on perceptions of evidence helpfulness, $F(1, 613) = 2.93, p = .09$, $\eta_p^2 = .005$. The evidence was viewed as more valid when it matched the defendant than when it did not. Similarly, there was a main effect of cross-examination type on perceptions of the validity of the evidence, $F(2, 613) = 21.85, p < .001$, $\eta_p^2 = .07$, but not on perceptions of evidence helpfulness, $F(2, 613) = 0.28, p = .76$, $\eta_p^2 = .001$. The evidence was viewed as less valid in the evidence-focused cross-examination condition than it was in the expert-focused ($p < .001$, $M_{\text{diff}} = -0.41$ [95% CI: -0.59, -0.23]) or non-substantive cross-examination conditions ($p < .001$, $M_{\text{diff}} = -0.48$ [95% CI: -0.67, -0.30]). There was a significant interaction between match type and cross-examination type on perceptions of evidence helpfulness, $F(2, 613) = 3.45, p = .03$, $\eta_p^2 = .01$; however, follow-up tests did not reveal significant differences between groups and there was not a significant interaction for evidence validity, $F(2, 613) = 2.14, p = .12$, $\eta_p^2 = .007$.

**Perceived understanding of the expert and evidence.** As additional measures of perceptions of the expert and evidence, I also examined participants’ perceived understanding of
the expert’s testimony and his forensic technique. An ANOVA revealed that there was a significant main effect of match type on participants’ ratings of how well they understood the expert’s testimony, $F(1, 619) = 5.20, p = .02, \eta_p^2 = .008$. Participants reported understanding the expert better when he testified that the evidence matched the defendant ($M = 6.29, SD = 0.92$) than when he testified that it did not match ($M = 6.10, SD = 1.09$). There was not a significant effect of evidence type, $F(3, 619) = 1.45, p = .23, \eta_p^2 = .007$, or cross-examination type, $F(2, 619) = 2.05, p = .13, \eta_p^2 = .007$, and no interactions reached significance (all $p$s > .2). However, there were main effects of evidence type, $F(3, 618) = 2.98, p = .03, \eta_p^2 = .01$, match type, $F(1, 618) = 12.02, p = .001, \eta_p^2 = .02$, and cross-examination type, $F(2, 618) = 3.33, p = .04, \eta_p^2 = .01$, on participants’ ratings of how well they understood the forensic technique used by the expert, although no interactions reached significance (all $p$s > .1). Participants reported understanding fingerprint analysis better than toolmark analysis ($p = .03, M_{diff} = 0.39$ [95% CI: 0.03, 0.75]); no other differences between evidence types were significant (all $p$s > .1) (see Table 7 for means). Participants reported understanding the forensic technique better when it resulted in a match than when it did not. They also reported understanding the evidence better when non-substantive cross-examination was used than when evidence-focused cross-examination was used ($p = .02, M_{diff} = 0.32$ [95% CI: 0.04, 0.60]); no other differences were significant (all $p$s > .3).

**Hypothesis 5:** Prior beliefs about the reliability of the relevant type of evidence will significantly predict perceptions of guilt and perceptions of the expert and evidence.

**Hypothesis 6:** Attitudes about science will not significantly predict perceptions of guilt, but will predict perceptions of the expert and the evidence.
I examined both specific and general beliefs together in larger models, so I tested hypotheses five and six at the same time. My hypotheses were only partially supported. There was very little effect of pretrial beliefs on perceptions of guilt: specific naïve beliefs predicted only verdict confidence, and only negative general views about science predicted probability estimates (although follow-ups did not always indicate significant relationships). As expected, pretrial beliefs did influence perceptions of the evidence. Specific naïve beliefs had the strongest impact, influencing both perceptions of the credibility of the expert and of the quality of the evidence, while positive and negative views about science had some influence on perceptions of evidence validity.

In all of the below models, I included evidence type, match type, and cross-examination type as categorical predictors or between-subjects factors and specific naïve beliefs, beliefs-in-science, and reservations-about-science as continuous predictors or covariates. However, so as not to be repetitive, results are not reported for manipulated IVs (evidence type, match type, and cross-examination type) unless they are different from the models which do not include pretrial beliefs. Results and follow-ups for these variables can be found in the sections above.

**Verdicts and verdict confidence.** A logistic regression model for verdict choice was significant, $\chi^2(20) = 246.11, p < .001, R^2_{\text{nag}} = .48$. However, neither specific naïve beliefs, Wald $\chi^2(1) = 2.74, p = .10, \text{OR} = 1.26 (95\% \text{ CI} = 0.96, 1.67)$, beliefs-in-science, $\chi^2(1) = 1.02, p = .31, \text{OR} = 1.07 (95\% \text{ CI} = 0.94, 1.21)$, nor reservations-about-science, $\chi^2(1) = 3.37, p = .07, \text{OR} = 1.14 (95\% \text{ CI} = 0.99, 1.30)$, significantly predicted verdicts.

Naïve pretrial beliefs did influence verdict confidence. An ANCOVA revealed a significant effect of specific naïve beliefs on verdict confidence, $F(1, 548) = 5.52, p = .02, \eta^2_p = .01$, and when pretrial beliefs were included in the model, there was no longer a significant effect
of evidence type on verdict confidence, $F(3, 548) = 0.77, p = .51, \eta_p^2 = .004$. Follow-up analysis indicated that there was a small positive association between specific naïve beliefs and verdict confidence, $r(576) = .15, p < .001$. Neither beliefs-in-science, $F(1, 548) = 1.36, p = .25, \eta_p^2 = .002$, nor reservations-about-science, $F(1, 548) = 0.002, p = .96, \eta_p^2 < .001$, influenced verdict confidence.

**Probability estimates.** When pretrial beliefs were included in the ANCOVA model for estimates of the probability that the defendant murdered the victim, there was a significant effect of reservations-about-science, $F(1, 547) = 6.06, p = .01, \eta_p^2 = .01$; however, there was not a significant effect of specific naïve beliefs, $F(1, 547) = 1.25, p = .27, \eta_p^2 = .002$, or beliefs-in-science, $F(1, 547) = 0.12, p = .72, \eta_p^2 < .001$. Follow-up analysis indicated that there was a weak and marginally significant association between reservations-about-science and murder probability ratings, $r(688) = .07, p = .06$.

For estimates of the probability that the defendant was the source of the evidence, there was again a significant effect of reservations-about-science, $F(1, 546) = 4.96, p = .003, \eta_p^2 = .009$; however, follow-up analysis indicated that the association between reservations-about-science and source probability ratings did not reach significance, $r(637) = .06, p = .13$. There was no significant effect of specific naïve beliefs, $F(1, 546) = 2.80, p = .01, \eta_p^2 = .005$, or beliefs-in-science, $F(1, 546) = 0.05, p = .82, \eta_p^2 < .001$.

**Perceptions of expert credibility.** When pretrial beliefs were included in the ANCOVA examining expert credibility, there was a significant effect of specific naïve beliefs, $F(1, 543) = 35.63, p < .001, \eta_p^2 = .06$, and no longer a significant main effect of evidence type, $F(3, 543) = 0.82, p = .49, \eta_p^2 = .004$. Follow-up analysis indicated that there was a moderate positive association between specific naïve beliefs and expert credibility ratings, $r(571) = .31, p < .001$. 

56
There was no significant effect of beliefs-in-science, $F(1, 543) = 2.52, p = .11, \eta_p^2 = .005$, or of reservations-about-science, $F(1, 543) = 1.07, p = .30, \eta_p^2 = .002$.

**Perceptions of evidence quality.** When pretrial beliefs were included in the model examining perceptions of the quality of the evidence, there were significant effects of specific naïve beliefs, Wilks’ $\lambda = .93, F(2, 542) = 22.08, p < .001, \eta_p^2 = .08$, and reservations-about-science, Wilks’ $\lambda = .99, F(2, 542) = 3.28, p = .04, \eta_p^2 = .01$, and a marginally significant effect of beliefs-in-science, Wilks’ $\lambda = .99, F(2, 542) = 2.80, p = .06, \eta_p^2 = .01$. Specific naïve beliefs significantly influenced both evidence validity ratings, $F(1, 543) = 44.3, p < .001, \eta_p^2 = .08$, and evidence helpfulness ratings, $F(1, 543) = 14.03, p < .001, \eta_p^2 = .03$. Follow-up analyses indicated that there were moderate positive associations between specific naïve beliefs and both evidence validity ratings, $r(573) = .38, p < .001$, and evidence helpfulness ratings, $r(574) = .24, p < .001$.

There was a significant effect of beliefs-in-science on evidence validity ratings, $F(1, 543) = 5.26, p = .02, \eta_p^2 = .01$, but no significant effect on evidence helpfulness ratings, $F(1, 543) = 3.06, p = .08, \eta_p^2 = .006$. Follow-up analysis indicated that there was a small but significant positive association between beliefs-in-science and evidence validity ratings, $r(636) = .18, p < .001$. Reservations-about-science had a significant effect on evidence validity ratings, $F(1, 543) = 4.43, p = .03, \eta_p^2 = .008$, but no effect on evidence helpfulness ratings, $F(1, 543) < .001, p = .98, \eta_p^2 < .001$. Follow-up analysis indicated that reservations-about-science had a small but significant negative association with evidence validity ratings, $r(635) = -.15, p < .001$.

**Perceived understanding of the expert and evidence.** I found significant effects of specific naïve beliefs, $F(1, 548) = 8.52, p = .004, \eta_p^2 = .02$, and reservations-about-science, $F(1, 548) = 6.41, p = .01, \eta_p^2 = .01$, on participants’ ratings of how well they understood the expert’s testimony; there was not a significant effect of beliefs-in-science, $F(1, 548) = 2.05, p = .15, \eta_p^2 = .
Follow-up analyses indicated that there was a small positive association between *specific naïve beliefs* and reported understanding of the expert, \( r(576) = .14, p = .001 \), and a small negative association between *reservations-about-science* and reported understanding of the expert, \( r(639) = -.14, p < .001 \).

Similarly, I found significant effects of *specific naïve beliefs*, \( F(1, 547) = 10.47, p = .001, \eta_p^2 = .02 \), and *reservations-about-science*, \( F(1, 547) = 5.35, p = .02, \eta_p^2 = .01 \), on participants’ ratings of how well they understood the forensic technique used by the expert; there was not a significant effect of *beliefs-in-science*, \( F(1, 547) = 1.39, p = .24, \eta_p^2 = .003 \), and the effect of cross-examination was no longer significant, \( F(1, 547) = 2.58, p = .08, \eta_p^2 = .009 \). Follow-up analyses indicated that there was again a small positive association between *specific naïve beliefs* and reported understanding of the forensic technique, \( r(575) = .13, p = .002 \), and a small negative association between *reservations-about-science* and reported understanding of the technique, \( r(638) = -.14, p < .001 \).

**Hypothesis 7: Pretrial beliefs will function as better predictors of perceptions of guilt and perceptions of the expert and evidence for high NC than low NC participants.**

My seventh hypothesis was only partially supported. Contrary to my expectations, pretrial beliefs predicted perceptions of guilt for low but not high NC participants. Pretrial beliefs predicted perceptions of the expert and evidence for all NC groups, but the relationship was strongest for high NC participants. I first examined whether pretrial beliefs differed for the three NC groups, and then included pretrial beliefs in logistic regression, ANCOVA, and MANCOVA models for each NC group.
Specific and general naïve beliefs. I first examined whether there were pretrial differences in beliefs regarding the specific evidence type participants read about or in general beliefs about science. I did not find any differences between NC groups in specific naïve beliefs, $F(2, 559) = 0.15, p = .86, \eta^2_p = .001$. There were, however, significant differences between groups in both beliefs-in-science, $F(2, 671) = 10.66, p < .001, \eta^2_p = .03$, and reservations-about-science, $F(2, 669) = 29.83, p < .001, \eta^2_p = .08$. High NC participants had higher beliefs-in-science ratings than either low ($p < .001, M_{diff} = 0.75 [95\% CI: 0.34, 1.16]$) or moderate NC participants ($p = .001, M_{diff} = 0.65 [95\% CI: 0.23, 1.07]$) (see Table 8 for means). Low and moderate NC participants did not differ ($p = .85, M_{diff} = 0.10 [95\% CI: -0.32, 0.52]$). Unsurprisingly then, low NC participants had higher reservations-about-science ratings than either high ($p < .001, M_{diff} = 1.24 [95\% CI: 0.85, 1.62]$) or moderate NC participants ($p = .04, M_{diff} = 0.40 [95\% CI: 0.01, 0.79]$). Moderate and high NC participants also differed significantly ($p = .001, M_{diff} = 0.84 [95\% CI: 0.45, 1.23]$).

Verdicts. When I included specific naïve beliefs, beliefs-in-science, and reservations-about-science as predictors in a logistic regression model and looked at each NC group separately, I found differences between NC groups in whether and what type of pretrial beliefs influenced verdicts. General beliefs predicted low NC verdicts, specific beliefs predicted moderate NC verdicts, and neither predicted high NC verdicts. For low NC participants, the logistic regression model was significant, $\chi^2(3) = 8.46, p = .03, R^2_{nag} = .06$, and both beliefs-in-science, Wald $\chi^2(1) = 4.43, p = .04, OR = 1.23 [95\% CI: 1.01, 1.48]$, and reservations-about-science, Wald $\chi^2(1) = 4.69, p = .03, OR = 1.25 [95\% CI: 1.02, 1.53]$, significantly predicted verdicts; there was no effect of specific naïve beliefs, Wald $\chi^2(1) = 2.28, p = .13, OR = 1.30 [95\% CI: 0.92, 1.84]$. However, although the directions of both associations were positive,
follow-up analyses did not reveal significant relationships between verdicts and either beliefs-in-science, \( r(231) = .11, p = .10 \), or reservations-about-science, \( r(230) = .11, p = .11 \). For moderate NC participants, the logistic regression model was marginally significant, \( \chi^2(3) = 7.42, p = .06 \), \( R^2_{\text{nag}} = .05 \), and specific naïve beliefs, Wald \( \chi^2(1) = 4.29, p = .04 \), OR = 1.51 (95% CI: 1.02, 2.23), significantly predicted verdict choice; neither beliefs-in-science, Wald \( \chi^2(1) = 1.65, p = .20 \), OR = 0.89 (95% CI: 0.74, 1.06), nor reservations-about-science, Wald \( \chi^2(1) = 0.23, p = .63 \), OR = 1.05 (95% CI: 0.87, 1.26), were significant predictors. Follow-up analysis indicated that there was a small positive association between specific naïve beliefs and verdicts, \( r(231) = .16, p = .03 \). For high NC participants, the logistic regression model was not significant, \( \chi^2(3) = 1.79, p = .62 \), \( R^2_{\text{nag}} = .01 \).

**Verdict confidence.** For verdict confidence, only low NC participants were influenced by pretrial beliefs. For low NC participants, there was a significant effect of specific naïve beliefs, \( F(1, 181) = 8.93, p = .003, \eta_p^2 = .05 \), on verdict confidence; there was not a significant effect of beliefs-in-science, \( F(1, 181) = 1.61, p = .21, \eta_p^2 = .009 \), or reservations about science, \( F(1, 181) = 2.57, p = .11, \eta_p^2 = .05 \). Follow-up analysis indicated that there was a moderate positive association between specific naïve beliefs and verdict confidence, \( r(184) = .21, p = .004 \). For moderate NC participants, there was no effect of specific naïve beliefs, \( F(1, 185) = 2.49, p = .12, \eta_p^2 = .01 \), beliefs-in-science, \( F(1, 185) = 0.65, p = .42, \eta_p^2 = .004 \), or reservations-about-science, \( F(1, 185) = 0.12, p = .73, \eta_p^2 = .001 \), on verdict confidence. For high NC participants, there was also no effect of specific naïve beliefs, \( F(1, 181) = 2.13, p = .15, \eta_p^2 = .01 \), beliefs-in-science, \( F(1, 181) = 2.14, p = .15, \eta_p^2 = .01 \), or reservations-about-science, \( F(1, 181) = 0.40, p = .53, \eta_p^2 = .002 \), on verdict confidence.
Probability estimates. For ratings of the probability that the defendant murdered the victim, low NC participants were again the only group influenced by pretrial beliefs. For low NC participants, there was a significant effect of reservations-about-science, $F(1, 180) = 9.01, p = .003, \eta_p^2 = .05$; there was no effect of specific naïve beliefs, $F(1, 180) = 1.89, p = .17, \eta_p^2 = .01$, or of beliefs-in-science, $F(1, 180) = 3.24, p = .07, \eta_p^2 = .02$. Follow-up analyses indicated that there was a small positive relationship between reservations-about-science and estimates of the probability that the defendant murdered the victim, $r(229) = .16, p = .02$. For moderate NC participants, there was no effect of specific naïve beliefs, $F(1, 185) = 0.84, p = .36, \eta_p^2 = .005$, beliefs-in-science, $F(1, 185) = 1.74, p = .19, \eta_p^2 = .009$, or reservations-about-science, $F(1, 185) = 2.12, p = .15, \eta_p^2 = .01$, on estimates of the probability that the defendant murdered the victim. For high NC participants, there was also no effect of specific naïve beliefs, $F(1, 181) = 0.30, p = .59, \eta_p^2 = .002$, beliefs-in-science, $F(1, 181) = 0.29, p = .59, \eta_p^2 = .002$, or reservations-about-science, $F(1, 181) = 0.39, p = .53, \eta_p^2 = .002$, on estimates of the probability that the defendant murdered the victim.

For ratings of the probability that the defendant was the source of the evidence, only moderate NC participants were influenced by pretrial beliefs, and even they were influenced very little. For low NC participants, there was no effect of specific naïve beliefs, $F(1, 180) = 0.30, p = .59, \eta_p^2 = .002$, beliefs-in-science, $F(1, 180) = 2.69, p = .10, \eta_p^2 = .02$, or reservations about science, $F(1, 180) = 2.48, p = .12, \eta_p^2 = .01$, on estimates of the probability that the defendant was the source of the evidence. For moderate NC participants there was a marginally significant effect of beliefs-in-science, $F(1, 185) = 3.50, p = .06, \eta_p^2 = .02$, on estimates of the probability that the defendant was the source of the evidence; there was no effect of specific naïve beliefs, $F(1, 185) = 2.01, p = .16, \eta_p^2 = .01$, or reservations-about-science, $F(1, 185) = 0.29, p = .59, \eta_p^2 = .02$.
Follow-up analysis indicated that there was a small negative association between beliefs-in-science and estimates of the probability that the defendant was the source of the evidence, $r(203) = -.16, p = .02$. For high NC participants, there was no effect of specific naïve beliefs, $F(1, 180) = 0.32, p = .57, \eta^2_p = .002$, beliefs-in-science, $F(1, 180) = 0.22, p = .64, \eta^2_p = .001$, or reservations-about-science, $F(1, 180) = 0.45, p = .50, \eta^2_p = .002$, on estimates of the probability that the defendant was the source of the evidence.

**Perceptions of expert credibility.** Pretrial beliefs about the type of evidence viewed influenced all three NC groups in their ratings of the credibility of the expert, although the relationship was strongest for high NC participants. For low NC participants, there was a significant effect of specific naïve beliefs, $F(1, 178) = 16.25, p < .001, \eta^2_p = .08$; there was not a significant effect of beliefs-in-science, $F(1, 178) = 0.62, p = .43, \eta^2_p = .003$, or reservations-about-science, $F(1, 178) = 0.24, p = .63, \eta^2_p = .001$. Follow-up analysis indicated a moderate positive association between specific naïve beliefs and ratings of expert credibility, $r(181) = .30, p < .001$. Similarly, for moderate NC participants, there was a significant effect of specific naïve beliefs, $F(1, 184) = 12.83, p < .001, \eta^2_p = .07$; there was not a significant effect of beliefs-in-science, $F(1, 184) = 2.67, p = .10, \eta^2_p = .01$, or reservations-about-science, $F(1, 184) = 2.23, p = .14, \eta^2_p = .01$. Follow-up analysis indicated a moderate positive association between specific naïve beliefs and ratings of expert credibility, $r(186) = .27, p < .001$. For high NC participants, there was also a significant effect of specific naïve beliefs, $F(1, 180) = 29.25, p < .001, \eta^2_p = .005$; there was not a significant effect of beliefs-in-science, $F(1, 180) = 0.00, p = .99, \eta^2_p < .001$, or reservations-about-science, $F(1, 180) = 0.20, p = .66, \eta^2_p = .001$. Follow-up analysis indicated a moderately strong positive association between specific naïve beliefs and ratings of expert credibility, $r(184) = .37, p < .001$. 
**Perceptions of evidence quality.** Pretrial beliefs also influenced perceptions of the evidence for all three NC groups, although again the strength of the association was greatest for high NC participants. For low NC participants, there was a significant effect of *specific naïve beliefs* on evidence quality ratings, Wilks’ $\lambda = .86$, $F(2, 178) = 14.80$, $p < .001$, $\eta_p^2 = .14$. There was no significant effect of *reservations-about-science*, Wilks’ $\lambda = .97$, $F(2, 178) = 2.61$, $p = .08$, $\eta_p^2 = .03$, or *beliefs-in-science*, Wilks’ $\lambda = .98$, $F(2, 178) = 2.10$, $p = .13$, $\eta_p^2 = .02$. *Specific naïve beliefs* had a significant effect on both ratings of evidence validity, $F(1, 179) = 27.07$, $p < .001$, $\eta_p^2 = .13$, and evidence helpfulness, $F(1, 179) = 14.91$, $p < .001$, $\eta_p^2 = .08$. Follow-up analyses indicated that there was a moderately strong positive relationship between *specific naïve beliefs* and evidence validity ratings, $r(182) = .38$, $p < .001$, and a moderate positive relationship between *specific naïve beliefs* and evidence helpfulness ratings, $r(184) = .27$, $p < .001$.

For moderate NC participants, there was a significant effect of *specific naïve beliefs*, Wilks’ $\lambda = .89$, $F(2, 182) = 10.98$, $p < .001$, $\eta_p^2 = .11$, on evidence ratings; there was no effect of *beliefs-in-science*, Wilks’ $\lambda = .99$, $F(2, 182) = 0.75$, $p = .48$, $\eta_p^2 = .008$, or *reservations-about-science*, Wilks’ $\lambda = .97$, $F(2, 182) = 0.19$, $p = .83$, $\eta_p^2 = .002$. *Specific naïve beliefs* had a significant effect on both ratings of evidence validity, $F(1, 183) = 21.35$, $p < .001$, $\eta_p^2 = .10$, and evidence helpfulness, $F(1, 183) = 5.84$, $p = .02$, $\eta_p^2 = .03$. Follow-up analyses indicated that there was a moderate positive relationship between *specific naïve beliefs* and evidence validity ratings, $r(186) = .33$, $p < .001$, and a small positive relationship between *specific naïve beliefs* and evidence helpfulness ratings, $r(186) = .16$, $p = .02$.

For high NC participants, there was a significant effect of *specific naïve beliefs*, Wilks’ $\lambda = .79$, $F(2, 179) = 24.46$, $p < .001$, $\eta_p^2 = .22$, on evidence ratings; there was no effect of *beliefs-in-science*, Wilks’ $\lambda = 1.00$, $F(2, 179) = 0.20$, $p = .82$, $\eta_p^2 = .002$, or *reservations-about-science*,
Wilks’ $\lambda = .98$, $F(2, 179) = 1.47, p = .23$, $\eta_p^2 = .02$. Specific naïve beliefs had a significant effect on both ratings of evidence validity, $F(1, 180) = 48.64, p < .001$, $\eta_p^2 = .14$, and evidence helpfulness, $F(1, 180) = 18.03, p < .001$, $\eta_p^2 = .04$. Follow-up analyses indicated that there was a strong positive relationship between specific naïve beliefs and evidence validity ratings, $r(185) = .45, p < .001$, and a moderate positive relationship between specific naïve beliefs and evidence helpfulness ratings, $r(184) = .29, p < .001$.

**Hypothesis 8:** Low NC participants will be equally influenced by both types of substantive cross-examination (expert-focused and evidence-focused), while high NC participants will differentiate between types of cross-examination such that expert-focused cross-examination will be more effective with more valid types of evidence, while evidence-focused cross-examination will be more effective with less valid types of evidence.

My eighth hypothesis was not supported: I found no differences between low and high NC participants in how they were influenced by the cross-examination manipulation. I did find some influence of NC more broadly: high NC participants were most influenced by the match manipulation and rated the expert as most credible and the evidence as highest in quality. They also gave the highest ratings of their understanding of the expert and the evidence. For the following analyses, all models included NC, evidence type, match type, and cross-examination type; Tukey tests were used for all post hoc analyses. To avoid repetition, only main effects of NC or interactions with NC are reported unless an effect or interaction was different from in the primary analyses.

**Verdicts and verdict confidence.** A logistic regression model for verdict choice including NC, evidence type, match type, cross-examination type and all two-way interactions
with NC, was significant, $\chi^2(20) = 243.30, p < .001, R^2_{\text{mag}} = .44$. However, neither NC, nor any two-way interactions with NC were significant (all $p$s > .4). For verdict confidence, an ANOVA revealed that there was a significant effect of NC, $F(2, 552) = 6.28, p = .002, \eta_p^2 = .02$, and a significant three-way interaction between NC, evidence type, and match type, $F(6, 552) = 3.47, p = .002, \eta_p^2 = .04$. Post hoc Tukey tests indicated that high NC participants tended to have higher confidence than low NC participants, although the difference did not reach significance ($p = .09, M_{\text{diff}} = 0.34 [95\% \text{ CI}: -0.04, 0.73]$). Moderate NC participants did not differ from either high ($p = .15, M_{\text{diff}} = -0.31 [95\% \text{ CI}: -0.70, 0.08]$) or low NC participants ($p = .98, M_{\text{diff}} = 0.03 [95\% \text{ CI}: -0.36, 0.42]$) (for means see Table 9). To investigate the three-way interaction, I conducted follow-up ANOVAs including evidence type and cross-examination type for each of the three NC groups which indicated that the interaction was significant only for high NC participants, $F(3, 202) = 3.82, p = .01, \eta_p^2 = .05$; it was not significant for moderate NC, $F(3, 197) = 1.54, p = .21, \eta_p^2 = .02$, or low NC participants, $F(3, 201) = 2.08, p = .10, \eta_p^2 = .03$. For high NC participants, when the evidence did not match the defendant, there was a significant effect of evidence type, $F(3, 103) = 3.83, p = .01, \eta_p^2 = .10$; when the evidence did match the defendant, the effect was not significant, $F(3, 99) = 2.00, p = .12, \eta_p^2 = .06$. In non-match conditions, verdict confidence was significantly higher in fingerprint ($M = 5.72, SD = 1.23$) than bitemark conditions ($M = 4.08, SD = 2.00; p = .01$); no other differences reached significance (all $p$s > .1).

**Probability estimates.** For estimates of the probability that the defendant murdered the victim, there was no main effect of NC, $F(2, 551) = 0.08, p = .92, \eta_p^2 < .001$, and no interactions with NC were significant (all $p$s > .4). For estimates of the probability that the defendant was the source of the evidence, there was again no main effect of NC, $F(2, 550) = 1.23, p = .30, \eta_p^2 = .004$, but there was a significant interaction between NC and match type, $F(2, 550) = 9.56, p <$
Although participants in all groups said it was more likely that the defendant was the source of the evidence when the evidence matched than when it did not, high NC participants took a non-match into consideration to a greater extent than moderate or low NC participants (see Table 10). It appeared that in general, the match manipulation had a greater impact on high NC participants than it did on low or moderate NC participants, whose ratings were almost identical.

**Perceptions of expert credibility.** For ratings of expert credibility, there was a significant main effect of NC, $F(2, 547) = 8.62, p < .001, \eta_{p}^2 = .03$, and a significant four-way interaction between NC, evidence type, match type, and cross-examination type, $F(2, 547) = 1.79, p = .048, \eta_{p}^2 = .04$. Post hoc Tukey tests indicated that high NC participants rated the expert as significantly more credible than low ($p < .001, M_{\text{diff}} = 0.38 [95\% \text{ CI: } 0.17, 0.60]$) or moderate NC participants ($p = .001, M_{\text{diff}} = 0.34 [95\% \text{ CI: } 0.12, 0.55]$) (for means see Table 9). Moderate NC participants did not differ from low NC participants ($p = .85, M_{\text{diff}} = 0.05 [95\% \text{ CI: } -0.17, 0.27]$). Follow-up ANOVAs for each NC group revealed that the three-way interaction between evidence type, match type, and cross-examination was significant for moderate NC participants only, $F(2, 180) = 3.40, p = .003, \eta_{p}^2 = .10$; it was not significant for either low, $F(2, 182) = 0.51, p = .80, \eta_{p}^2 = .02$, or high NC participants, $F(2, 185) = 1.36, p = .24, \eta_{p}^2 = .04$. For moderate NC participants, when the expert testified that the evidence did not match the defendant, there were no significant differences between evidence conditions for any type of cross-examination (all $p$s > .08). When the expert testified that the evidence matched the defendant, when expert-focused cross-examination was used, the DNA expert ($M = 6.69, SD = 0.36$) was rated as significantly more credible than the toolmark ($M = 5.49, SD = 0.92; p = .02, M_{\text{diff}} = 1.20 [95\% \text{ CI: } 0.19, 2.21]$) or bitemark expert ($M = 5.28, SD = 1.14; p = .02, M_{\text{diff}} = 1.41$).
the fingerprint expert \((M = 6.40, SD = 0.45)\) was rated as significantly more credible than the bitemark expert \((p = .045, M_{diff} = 1.12 [95\% CI: 0.02, 2.22])\) and marginally more credible than the toolmark expert \((p = .05, M_{diff} = 0.91 [95\% CI: -0.01, 1.83])\). No other differences reached significance (all \(p s > .4\)).

**Perceptions of evidence quality.** For perceptions of the evidence, a MANOVA revealed a significant main effect of NC, Wilks’ \(\lambda = .96\), \(F(2, 1090) = 6.22, p < .001, \eta_p^2 = .02\); however, no interactions with NC were significant. Univariate tests indicated that both ratings of evidence validity, \(F(2, 615) = 10.10, p < .001, \eta_p^2 = .03\), and evidence helpfulness, \(F(2, 615) = 3.18, p = .04, \eta_p^2 = .01\), differed significantly based on NC group. For evidence validity, post hoc Tukey tests indicated that high NC participants gave significantly higher validity ratings than low \((p < .001, M_{diff} = 0.35 [95\% CI: 0.16, 0.54])\) or moderate NC participants \((p = .002, M_{diff} = 0.29 [95\% CI: 0.09, 0.48])\) (for means see Table 9). Moderate NC participants did not differ from low NC participants \((p = .70, M_{diff} = 0.06 [95\% CI: -0.13, 0.26])\). Post hoc Tukey tests indicated that high NC participants gave significantly higher evidence helpfulness ratings than low NC participants \((p = .04, M_{diff} = 0.30 [95\% CI: 0.01, 0.59])\). Moderate NC participants did not differ from either high \((p = .29, M_{diff} = -0.13 [95\% CI: -0.43, 0.16])\) or low NC participants \((p = .58, M_{diff} = 0.17 [95\% CI: -0.12, 0.46])\).

**Perceived understanding of the expert and evidence.** For reported understanding of the expert and evidence, ANOVAs revealed significant main effects of NC on participants’ ratings of how well they understood the expert’s testimony, \(F(2, 552) = 17.99, p < .001, \eta_p^2 = .06\), and the forensic technique upon which it was based, \(F(2, 551) = 12.48, p < .001, \eta_p^2 = .04\), but no interactions with NC reached significance (all \(p s > .2\)). Additionally, the main effect of cross-examination type on perceived understanding of the expert’s forensic technique was no longer
significant, $F(2, 551) = 2.65, p = .07, \eta^2_p = .01$. High NC participants said that they understood both the expert and the evidence better than low NC (expert $p < .001, M_{\text{diff}} = 0.56$ [95% CI: 0.33, 0.79]; evidence $p < .001, M_{\text{diff}} = 0.66$ [95% CI: 0.37, 0.94]) or moderate NC participants (expert $p < .001, M_{\text{diff}} = 0.40$ [95% CI: 0.17, 0.63]; evidence $p = .02, M_{\text{diff}} = 0.32$ [95% CI: 0.04, 0.61]) (for means see Table 9). Moderate NC participants said that they understood the forensic technique used by the expert significantly better than low NC participants ($p = .02, M_{\text{diff}} = 0.33$ [95% CI: 0.04, 0.62]), but their ratings of how well they understood the expert’s testimony did not differ ($p = .25, M_{\text{diff}} = 0.16$ [95% CI: -0.07, 0.38]).
CHAPTER 9: STUDY ONE DISCUSSION

Generally, participants expressed positive views of science and technology, with a high percentage of participants agreeing with each of the individual questions regarding the benefits of science. Some participants also expressed negative views about science, although to a much lesser degree; fewer than a quarter of participants agreed or strongly agreed with questions reflecting reservations about science. Overall, my sample had similarly positive views about science to prior samples (Hans et al., 2011; J. D. Miller, 2004), although a smaller percentage than in those samples expressed reservations about science. Online participation suggests at least some level of comfort with technology, so it is perhaps not surprising that my online participants expressed fewer reservations regarding science and technology than the prior, off-line samples.

Consistent with previous research (Hans et al., 2011; Lieberman et al., 2008), participants viewed forensic science more positively than other types of evidence. Reliability ratings were significantly higher for all four types of forensic science that I examined than they were for any other type of evidence, suggesting that participants tended to overestimate the validity of some of these types of evidence, in particular that of bitemark and toolmark evidence. Although not as apparent from reliability ratings, they also appeared to overestimate the validity of fingerprint evidence, given that post-trial ratings and judgments in fingerprint conditions suggested that it was given similar weight to DNA evidence. As expected, DNA was rated as more reliable than all other types of evidence followed by fingerprint evidence. However, bitemark evidence was rated as more reliable than toolmark evidence, which was unexpected and is not consistent with the actual state of the science (Harris, 2012; NAS, 2009). It is possible that the lower ratings were due at least in part to participants being less familiar with toolmark evidence than the other types of evidence (as indicated by a much larger number reporting never to have heard of
toolmark evidence than any other type of evidence examined)—generally, people tend to view things with which they are more familiar more positively than those with which they are not familiar (Zajonc, 1968). However, it should also be noted that pretrial beliefs were assessed using only one question that asked about reliability using a 5-point scale. It is possible that participants might have responded differently if they had been presented with a more sensitive measure that directly assessed perceptions of scientific quality. Thus, in Studies 2 and 3, I asked participants about reliability and about scientific quality and used a much larger scale to give me greater power to uncover potential differences in beliefs about evidence types.

Specific pretrial beliefs regarding the type of evidence participants read about in their assigned transcript did not significantly influence participants’ verdict choices, or their estimates of the likelihood that the defendant murdered the victim or was the source of the evidence. However, pretrial beliefs were associated with participants’ confidence in their verdict choice: participants were more confident in their verdicts when they believed going into the trial that the type of evidence they read about was more reliable. This finding may indicate that the more that participants thought the scientific trial evidence was valid, the stronger the basis they felt they had for their verdict choice, a relationship that makes intuitive sense.

Not surprisingly, pretrial beliefs regarding the type of evidence participants read about were associated with participants’ ratings of the expert and the evidence. The more participants believed that the type of evidence was valid prior to reading the transcripts, the more credible they found the expert and the more valid and helpful they found the evidence. Valid evidence should be weighed more strongly than invalid evidence; thus, these results suggest that if participants possessed more accurate pretrial beliefs regarding the validity of forensic science, they might be better equipped to weigh such evidence appropriately at trial. However, given that
these beliefs did not significantly influence verdicts, it is not clear whether improving the accuracy of pretrial beliefs would be sufficient to influence jurors’ ultimate judgments; I sought clarification of this issue in my second and third studies.

Interestingly, negative views about science in general were more predictive of my dependent measures than positive views about science. Views about the benefits of science did not predict any of my measures of perceptions of guilt, although there was a positive association between beliefs-in-science and perceptions of the validity of the evidence. However, views reflecting reservations about science were predictive of verdict and both probability estimates, although the association was small (and was not significant for estimates of the probability that the defendant was the source of the evidence). Although one might expect that in a trial where the primary evidence was scientific in nature, reservations regarding science and technology might be negatively associated with perceptions of guilt, in all cases I found positive relationships. By contrast, I found that the correlation between reservations-about-science and perceptions of evidence validity was negative, as expected. These findings suggest the possibility that people who are more reluctant to embrace science—or more specifically those who feel that it has eroded or come at the expense of faith—may be generally more conviction prone, even if they have less faith specifically in scientific evidence. It is possible that such views are connected to greater religious affiliation or to a stronger attachment to conventional values. There is evidence that some types of religious beliefs may be associated with greater punitiveness (M. K. Miller, Singer, & Jehle, 2008). Further, adherence to conventional values, as part of the more general trait of authoritarianism, has also been associated with punitiveness (Narby et al., 1993).
As one might expect based on their inflated pretrial ratings of forensic science reliability, participants tended to give more weight than deserved to less valid types of forensic science. Even bitemark evidence, the least valid of those included in the study, had a significant influence on verdicts, with just over half of participants (51.3%) finding the defendant guilty when presented with a bitemark match. The guilty verdict rate for toolmark evidence, also of very questionable validity (NAS, 2009) was almost identical (51.4%). Indeed, the percentage of guilty verdicts in these conditions was almost three times that found in the control group (17.6%). The NAS Report was quite critical of both toolmark and bitemark evidence, and concluded that at present, it is not clear whether a bitemark match has any probative value. Thus, a finding that an expert testifying about a toolmark or bitemark match made participants so much more likely to find the defendant guilty suggests that people give too much weight to disciplines that are not supported by the scientific community—at least when such evidence inculpates the defendant.

Consistent with previous research (Garrett & Mitchell, 2013; McAllister & Bregman, 1986), a match had a greater influence than a non-match. When an expert testified that evidence matched the defendant, the guilty verdict rate was significantly higher than in the control group, while when he testified that it did not match, verdicts did not differ from the control group. A match also had a greater influence than a non-match on perceptions of the likelihood that the defendant murdered the victim; however, contrary to my expectations, both were significantly different than the control group. The defendant was rated as significantly more likely to have murdered the victim in match conditions and significantly less likely in non-match conditions.

Unexpectedly, whether the evidence was presented as a match to the defendant also influenced perceptions of the expert and the evidence. The expert was rated as more credible and the evidence as more valid when the expert testified that it matched the defendant than when he
said that it did not match. While one might presume that this difference could be due to a match being seen as more useful to the case, perceived utility cannot explain this finding given that helpfulness ratings did not differ between match and non-match conditions. Instead, it appears that forensic evidence and experts are viewed as higher in quality when such evidence ties the defendant to a case than when it does not. The reasoning for this difference is not clear. It is possible that despite the content of the expert’s testimony, people believe that a non-match is actually inconclusive or, worse, that on some level people expect forensic evidence to match if it is presented and presume that it is a failing on the part of the expert or the type of analysis when it does not. An alternative, and equally troubling, explanation is that the differences in perceptions related to the side presenting the defense. The match was presented by the prosecution, while the non-match was presented by the defense. It is possible that a defense expert—and his resulting analyses—are seen as less credible than a prosecution expert. Previous research suggests that the underutilization of a non-match in terms of probative value occurs regardless of whether the non-match is presented by the prosecution or defense: Garrett and Mitchell’s (2013) non-match was presented by the prosecution and they found that perceptions of guilt were similar to their control group. They used only brief case summaries, however, and did not access perceptions of the expert, so it is not clear whether their participants also perceived a non-match as of lesser quality. That said, given that a non-match is more likely to be presented by the defense, it is likely that in many cases the result would be the same regardless of the cause of the difference: more negative perceptions of evidence presented by the defense. Thus differences in perceived credibility could be quite prejudicial to defendants and suggests that defense attorneys may need to work harder than prosecutors to convey the quality of their forensic evidence. Alternatively, they may need to find a more effective way to convey to jurors
the significance of a match versus a non-match, so that a non-match can be understood as a conclusion in its own right.

There were generally no differences between evidence type groups when the expert testified that the evidence did not match the defendant. As expected based on prior research (Garrett & Mitchell, 2013; Lieberman et al., 2008; McAllister & Bregman, 1986), for the most part participants differentiated between evidence types only when the expert testified that the evidence matched the defendant. When the expert testified that the evidence matched the defendant, verdicts and estimates of both the probability that the defendant murdered the victim and that he was the source of the evidence differed based on evidence type. It appeared that participants did generally have some sense that some types of evidence should receive more weight than others. However, participants seemed to weigh fingerprint evidence similarly to DNA evidence, and bitemark evidence similarly to toolmark evidence (in fact, no differences between the latter two were found on any post-trial measure), which suggests that they are not totally aware of the deficiencies of non-DNA evidence or of differences between the less valid types of forensic science.

Participants were more persuaded by matching DNA evidence than matching toolmark or bitemark evidence; the percentage of guilty verdicts and both types of probability estimates were higher in DNA conditions than in either toolmark or bitemark conditions. However, neither verdicts nor probability ratings for participants in DNA and fingerprint conditions differed significantly. Further, when the evidence matched the defendant, verdicts and estimates of the probability that the defendant was the source of the evidence were also significantly higher in fingerprint conditions than they were in toolmark or bitemark conditions; estimates of the probability that the defendant murdered the victim did not differ significantly.
Similarly, participants’ ratings of the fingerprint expert’s credibility did not differ from those of the DNA expert, and both were rated as significantly more credible than the toolmark or bitemark expert. However, it appears that pretrial beliefs about specific evidence types may mediate the relationship between the type of evidence viewed and perceptions of the credibility of the expert: when specific naïve beliefs were accounted for, evidence type no longer significantly predicted perceptions of expert credibility. Participants did differentiate between DNA and fingerprint evidence in their ratings of evidence validity. Participants presented with DNA evidence rated the evidence as more valid than participants presented with fingerprint, toolmark, or bitemark evidence, while fingerprint evidence was rated as more valid than both toolmark and bitemark evidence. However, DNA and fingerprint evidence were rated as equally helpful and both as more helpful than bitemark evidence. DNA evidence was also rated as more helpful than toolmark evidence. Given that participants rated DNA evidence as both higher in reliability than fingerprint evidence prior to trial and higher in validity post-trial, it is not clear why they did not differentiate between the two in weighing the evidence to decide on a verdict or in estimating the likelihood that the defendant murdered the victim or was the source of the evidence. It is also not clear why there were no differences in perceptions of the credibility of the expert or the helpfulness of the evidence.

Cross-examination type did not influence verdicts or participants’ perceptions of the likelihood that the defendant murdered the victim. It did influence participants’ perceptions of the likelihood that the defendant was the source of the evidence, an indirect measure of perceived strength of the forensic evidence, as well as perceptions of the credibility of the expert and the validity of the evidence. Contrary to what I expected based on Lieberman et al.’s (2008) results, the effectiveness of evidence-focused and expert-focused cross-examination did not vary.
depending on the validity of the type of evidence presented. Instead, the evidence-focused cross-examination was generally the most effective type of cross-examination across all evidence types, significantly reducing the impact of expert testimony in both match and non-match conditions. When the evidence matched the defendant, estimates of the probability that the defendant was the source of the evidence were the lowest when evidence-focused cross-examination was used, although the difference between evidence-focused and expert-focused cross-examination did not reach significance. When the evidence did not match the defendant, estimates were highest when evidence-focused cross-examination was used. Evidence-focused cross-examination was also the most effective in reducing perceptions of both expert credibility and evidence validity. Expert-focused cross-examination also had some impact, however, significantly reducing expert credibility ratings compared to non-substantive cross-examination.

In contrast to McQuiston-Surrett and Saks’s (2009) results, the inclusion of substantive cross-examination did not make participants feel that they understood the evidence better. Cross-examination did not influence participants’ reported understanding of the expert’s testimony, and participants actually felt that they understood the evidence less when they heard evidence-focused cross-examination than non-substantive cross-examination. However, unlike McQuiston-Surrett and Saks, I did not have a condition in which there was no cross-examination. Thus it is possible that even non-substantive cross-examination made participants feel they understood the evidence better than they would have had no cross-examination been presented.

While NC was related to several of my dependent measures, I did not find all of my expected interactions. In general, there were few patterns for moderate NC participants, who were sometimes similar to low NC participants, sometimes to high NC participants, and sometimes to neither. However, there were some consistent differences between high and low
NC participants, which seemed to be partially based on pretrial beliefs. While there were no differences in specific naïve beliefs, high NC participants had the most positive views of science in general, while low NC participants had the most reservations about science. It is possible that these differences are related to differences in orientation toward analytical thought. Given that people high in NC have a greater preference for analytical thought (Cacioppo et al., 1983), science may generally be more appealing to them than people low in NC.

Although previous research found that people high in NC were more influenced by information presented prior to trial evidence (Haugtvedt & Petty, 1992; Kassin et al., 1990), I did not find that high NC jurors were always more influenced by their pretrial beliefs. That said, I did find differences in when pretrial beliefs tended to influence the different NC groups. Low NC participants’ perceptions of guilt were most influenced by pretrial beliefs: general beliefs about science predicted verdicts and estimates of the probability that the defendant murdered the victim, while specific naïve beliefs influenced verdict confidence. By contrast, pretrial beliefs did not influence high NC participants’ perceptions of guilt. However, both high and low NC participants’ specific naïve beliefs influenced their perceptions of expert credibility and evidence quality, although the relationship in both cases was stronger for high NC participants. High NC participants also generally had more positive views of the forensic evidence in the transcript: they rated the expert as more credible and the evidence as more valid than low or moderate NC participants. They also rated the evidence as more helpful than low NC participants. Further, they said that they understood both the evidence and the expert better than low or moderate NC participants.

Contrary to my expectations, there were few differences in the manner in which evidence, match, and cross-examination type influenced the different NC groups. In particular,
there were no differences in the manner in which low and high NC participants were influenced by the different forms of cross-examination. High NC participants were more influenced by the other experimental manipulations in the extent to which they agreed with the expert’s conclusions, however: high NC participants were more influenced by whether or not the expert testified that the evidence matched the defendant in their ratings of the probability that the defendant was the source of the evidence. These results are consistent with the findings of Garrett and Mitchell (2013), who found that more numerate participants found fingerprint evidence more convincing than less numerate participants. It is possible that in part high NC participants’ greater perceived understanding could account for their more positive perceptions of the expert and evidence and their greater adherence to the expert’s conclusions. This explanation is consistent with research suggesting that we are more likely to automatically accept information that we are able to understand (Gilbert et al., 1990, 1993).

Consistent with prior research (Kovera et al., 1999; Lieberman et al., 2008; McQuiston-Surrett & Saks, 2009), cross-examination did not influence verdicts, nor did it influence estimates of the probability that the defendant murdered the victim. It is possible that cross-examination and other traditional legal safeguards may not always be effective in part because they follow the presentation of the evidence. If jurors have already formed opinions about the evidence based on their prior beliefs and what they have heard during direct examination, they may interpret the cross-examination in light of what they already believe to be true (Darley & Gross; 1983; Lord et al., 1979; Nickerson, 1998; Saks et al., 2003). Thus, if they already believe that forensic evidence is valid, and then an expert testifies that he has found a match between the suspect and the evidence using methods that he describes as reliable and valid, jurors may believe that there is a quite strong likelihood that the defendant is the source of the evidence and
therefore linked to the crime. If their opinion is strong, it may be difficult to counter after the fact by attempting to cast doubt on the validity of the science. However, if the juror enters the trial believing that forensic evidence is not valid, they may be more skeptical of a forensic expert’s claims and more cautious in accepting their conclusions even during direct testimony, in which case cross-examination or judicial instruction may be more effective in attacking the content of the direct testimony. Thus, in Studies 2 and 3, I considered ways of inducing an appropriate level of skepticism in potential jurors.
CHAPTER 10: PART II: THE INFLUENCE OF MEDIA-INFORMED BELIEFS

One possible way of inducing skepticism about forensic science is to introduce media reports prior to trial. Generally, the media is a powerful tool for disseminating information and for shaping people’s perceptions and attitudes (M. K. Johnson, 2007; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). Given the far-reaching power of the media, it may serve as an effective forum for correcting the public’s misconceptions about the validity of the forensic sciences. Further, the extensive body of research on pretrial publicity indicates that news reports viewed prior to a trial can influence jurors’ perceptions and judgments following the presentation of trial evidence (Spano, Groscup, & Penrod, 2011; Steblay, Besirevic, Fulero, & Jimenez-Lorente, 1999).

Pretrial Publicity

Research on pretrial publicity (PTP) has demonstrated that the media can have a lasting effect on jurors, influencing both their judgments and their perceptions of trial evidence (Spano et al., 2011; Steblay et al., 1999). For example, jurors are more likely to view the defendant as guilty when they are exposed to media reporting the defendant’s negative character (Otto, Penrod, & Dexter, 1994) or prior criminal record (Hvistendahl, 1979; Otto et al., 1994), or to reports that the defendant confessed (Tans & Chaffee, 1966; Padawer-Singer & Barton, 1975), or that his blood was found on the victim’s shirt (Shaw & Skolnick, 2004). Importantly, jurors may not realize that their perceptions are affected by PTP. Indeed, research suggests that the media can influence juror verdicts even when jurors claim that they will be able to remain impartial (Moran & Cutler, 1991). Further, as with other types of inadmissible evidence, judicial instructions to disregard PTP can backfire and result in greater effects of PTP on juror perceptions of trial evidence (Kramer, Kerr, & Carol, 1990; Steblay et al., 1999).
PTP need not appeal to jurors’ emotions in order to exert its influence: both factual and emotional PTP have been found to influence jurors (Spano et al., 2011). In a study involving the adjudication of a hypothetical armed robbery, Kramer and Kerr (1989) compared the effects of emotional to factual PTP on mock jurors viewing long and short armed robbery trials (100 versus 10 minutes). The emotional PTP described an unrelated hit-and-run likely committed by the defendant that resulted in the death of a young child, while the factual PTP described incriminating evidence in the armed robbery case, and the defendant’s extensive criminal background. They found that both types of PTP increased the percentage of guilty verdicts and the length of suggested sentences following both short and long trials. Wilson and Bornstein (1998) compared the effects of emotional and factual PTP preceding a homicide trial. They found that both types of PTP increased the percentage of mock jurors choosing murder over manslaughter as well as perceptions of guilt using a more continuous measure, with no significant differences between the two types of PTP.

Of particular importance to the current project, PTP need not be specific to a case in order to influence jurors: general PTP has also been found to affect subsequent juror judgments (Greene, 1990; Spano et al., 2011). In the first study to report on general PTP, Greene and Loftus (1984) were conducting a mock jury study which included the testimony of an eyewitness, and noticed that their conviction rates dropped unexpectedly partway through data collection. In attempting to determine why, they discovered that the story of a man who had been wrongfully convicted due to a misidentification and subsequently exonerated had been prominently featured in the news. Thus, they conducted two studies examining whether wrongful conviction media influenced conviction rates in eyewitness trials. To do so, they compared conviction rates during the period that a wrongful conviction due to misidentification was featured in the news to rates
when the story was no longer present in the media; in the second study, they made a more concerted effort to determine who had read about the misidentification. In both studies, they found that conviction rates were lower when the wrongful conviction story was featured in the media than it was when the story was no longer present.

The initial study involved responses to naturally-occurring media and did not directly control participants’ exposure to the PTP prior to the experiment; thus Greene and Wade (1988) conducted a follow-up study in which general PTP was experimentally manipulated. Their results confirmed the effects of the general PTP: mock jurors were more likely to find the defendant guilty when exposed to a news report about an (unrelated) heinous crime, but less likely to find the defendant guilty when they had been exposed to a report about a wrongful conviction. In a second study, they varied whether the report described a case similar or dissimilar to the case in the trial given to participants. Although both types of reports affected jurors’ decisions, the influence of the similar story was greater.

If, as demonstrated by the research on PTP, exposure to even general media prior to a trial can influence post-trial perceptions and judgments of jurors, media reports that demonstrate the deficiencies of forensic science may make jurors’ perceptions of such evidence more negative, but, critically, may also reduce their reliance on such evidence if it is presented at trial. If so, there may be ways to present such information in media reports in order to make those reports more effective. If we view incorrect beliefs about forensic science as a form of misinformation, it is possible that some of the tools that researchers have used to correct misinformation in other contexts could also be used to correct misconceptions about the forensic sciences.
Correcting Misinformation

As summarized by Lewandowky et al. (2012), corrections to misinformation are most effective when they are able to overcome several of the following problems: simple retractions may not be sufficient to correct misinformation; any correction that repeats the misinformation—even if alongside a refutation—serves to reinforce the misinformation by increasing its familiarity; complicated or extensive corrections may be less cognitively appealing than simple misinformation; and corrections that threaten strongly-held worldviews or ideological views can actually strengthen these initial beliefs. Numerous researchers have shown that simple retractions cannot correct misinformation: even when the correction is both believed and remembered, people continue to believe the misinformation and to make inferences on its basis (e.g., Ecker, Lewandowsky, & Apai, 2011; Gilbert et al., 1990, 1993; H. M. Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). However, if these retractions are accompanied by an alternative account that can replace the misinformation, the retractions may be more successful in correcting it. For example, H. M. Johnson and Seifert (1994) found that people were more likely to believe a correction to misinformation regarding the cause of a warehouse fire (a closet filled with cans of oil paint and pressurized gas cylinders) when they were provided with an alternative cause to the fire (evidence of arson) than when the original information was simply retracted (the closet was empty). Similarly, they found that misinformation regarding the identity of a burglar (it may have been the victim’s son) was more likely to be rejected when an alternate suspect was given than when the original information was simply retracted. A follow-up study (Seifert, 2002) suggested that providing an explanation for the original misinformation (e.g., the closet was usually filled with paint and gas cylinders but was empty because a truckers’ strike had delayed
delivery) was also successful in reducing reliance on the misinformation, though not as successful as providing an alternative account.

Casting suspicion on the source of the misinformation or the source’s motives can be an effective method of discrediting the misinformation and enhancing the effectiveness of the correction (Lewandowsky et al., 2012). Casting suspicion may be effective both because information from a non-credible source is less likely to be believed, and because it provides a form of explanation for the misinformation. The effect of casting suspicion has also been investigated in a juror context and has been found to help alleviate the biasing effects of negative PTP on juror perceptions and judgments. Fein, McClosky, and Tomlinson (1997) had mock jurors read a news report which contained both negative character information and specific evidence tying the defendant to the crime. Half of the jurors also read a news report in which the defense attorney attacked the motives of the media, suggesting that they twist the facts in order to increase sales and even that the information may have been planted by the district attorney. Participants who read only the negative PTP were more likely to find the defendant guilty than participants who had not read negative PTP; however, the verdicts of participants who read the additional report casting suspicion on the negative PTP did not differ from verdicts of participants who had not read negative PTP. Thus, when the defense attorney was able to cast suspicion on the motives of the media presenting the negative PTP, the negative PTP no longer influenced jurors. A similar effect may be found on information received during the trial: if an attorney is able to cast suspicion on an expert’s credentials or motivation via cross-examination, it may reduce the impact of his testimony (e.g., Cooper & Neuhaus, 2000; Lieberman et al., 2008). Indeed, in Study 1, expert-focused cross-examination, which included questioning the
expert’s motivation, did have some influence on participants’ perceptions, although I did not find it to be as effective as attacking the forensic technique itself.

Repetition of a correction may also be an effective method of correcting the misinformation, particularly when the misinformation has been provided on multiple occasions (Ecker, Lewandowsky, Swire, & Chang, 2011). Repetition may also backfire, however, for several reasons: because the misinformation may be repeated—and thereby strengthened—alongside the correction, because repeating the correction too many times may actually cause suspicion, and because the correction may begin to appear more complex than the misinformation (Lewandowsky et al., 2012). People generally prefer simple explanations to more complicated explanations and simple explanations are seen as highly probable. For example, Lombrozo (2007) conducted a series of experiments examining people’s preferences for different types of explanations. She found that simplicity was actually preferred over probability when selecting explanations. In other words, people tended to choose simple explanations even when more complex explanations were more likely. Quite disproportionately high relative probabilities were necessary in order for participants to select a complex over a simple explanation. In fact, even when a simple explanation was 10 times less likely than a more complicated explanation, 41% of participants continued to choose the simple explanation. Thus, corrections to misinformation that are simpler than the misinformation may be more likely to be successful than more complex corrections.

When misconceptions are tied to a strongly held worldview or ideological belief, they can be particularly resistant to change (Lewandowsky et al., 2012). Indeed, because people are motivated to protect their beliefs, attempts to correct misinformation that is supported by their beliefs can result in a backfire effect wherein belief in the misinformation is actually
strengthened (Hart & Nisbet, 2012; Lewandowsky et al., 2012; Nyhan & Reifler, 2010; Nyhan, Reifler, Richey, & Freed, 2014). For example, Nyhan et al. (2014) found that public health campaigns correcting the misconception that the MMR vaccine causes autism actually decreased parents’ intentions to vaccinate future children, despite the fact that it also decreased their stated belief in the fact that the vaccine causes autism. Such effects might even generalize beyond the specific belief attacked: Munro (2010) found that rather than accept scientific findings that are inconsistent with their beliefs, people may actually discount scientific research more broadly. When people were confronted with belief-disconfirming science, they were not only more likely to believe that science was incapable of studying the particular phenomenon in question, but were also more likely to believe that science was incapable of studying other, unrelated phenomena. Thus, when the misinformation is related to beliefs, the correction may need to be specifically designed or worded so as to make it more consistent with or to affirm some aspect of the previously held belief (Hardisty, Johnson, & Weber, 2010; Lewandowsky et al., 2012).

It is possible that the use of a personal narrative illustrating the misconception may be more effective than attacking the facts underlying the erroneous belief, especially when misconceptions are strongly held or tied to worldviews. A story can be an extraordinarily persuasive tool, one that can be more effective than more traditional, fact-based methods of persuasion (Green & Brock, 2005; Slater, 2002). Further, if people become involved in a story, they may be less likely to produce counterarguments to the message contained within the story (Green, 2008; Slater, 2002), which could make stories particularly well suited for correcting misinformation. Even the inclusion of brief narratives in otherwise informational reports can increase their persuasiveness and reduce the production of counterarguments (Slater & Rouner, 1996).
Similar results have been found in the courtroom: anecdotal or descriptive clinical evidence has been found to be more persuasive than experimental or statistical evidence (e.g., Bornstein, 2004; Krauss & Sales, 2001). Further, one of the most prominent models of judicial decision-making is the story model, which posits that jurors organize trial information in narrative form, integrating evidence into the story that they have constructed (Pennington & Hastie, 1986, 1992). Thus, use of a personal narrative that highlights the deficiencies of forensic science might be an effective manner of changing beliefs about its validity. A narrative about an instance in which forensic science was wrong—where it identified the wrong person—may be particularly effective. In addition to providing the more general benefits of storytelling, it may make it easier for jurors to believe that a match may not be valid if they can easily bring to mind a scenario in which a match was shown to be invalid (Koehler & Macchi, 2004; Tversky & Kahneman, 1974), particularly when the serious consequences of an invalid match are also apparent. Further, if people have become involved in a narrative, it may also make them more receptive to factual information that follows the story. Thus, it is possible that supplementing a story with factual information might be a particularly powerful way to correct erroneous beliefs.
CHAPTER 11: STUDIES TWO AND THREE OVERVIEW

In Studies 2 and 3, I investigated the effectiveness of different types of factual and story-based news reports in influencing mock jurors’ pretrial beliefs about invalid forensic science and their subsequent perceptions of the evidence and judgments in a criminal trial. In Study 2, I focused on how to most effectively communicate factual, scientific findings in the media with the goal of correcting misconceptions about forensic evidence. Participants either read a set of criminal justice articles unrelated to forensic science, or also read one of four target articles. I focused on bitemark evidence in the target articles, as it received the most criticism out of the four evidence types examined in Study 1. The four articles relayed information from the NAS Report’s findings about the reliability of bitemark evidence. Based on prior research on correcting misinformation (see Lewandowsky et al., 2012), I varied the structure, framing, and/or content of the target articles to test which variation had the greatest impact on people’s beliefs about the evidence and their subsequent perceptions and judgments. Specifically, I compared the effectiveness of: 1) a detailed, complex report; 2) a simple report; 3) a report which affirmed the readers’ beliefs in the criminal justice system alongside the critical information; and 4) a report which supplemented the critical information with specific information about how bitemark evidence should be evaluated, offering the reader an alternative to their current method of evaluating evidence. After reading their assigned articles, participants rated the validity of different types of evidence—including bitemark evidence—and read a trial transcript containing either bitemark evidence or toolmark evidence (selected because toolmark and bitemark evidence were rated similarly in Study 1).

In Study 3, I examined whether a narrative report was more effective than a factual report in changing people’s beliefs about forensic science and their perceptions and judgments in a
subsequent criminal trial. I was also interested in generalizability, and whether reading a narrative report rather than a factual report would be more likely to broadly influence perceptions of forensic evidence. Thus in Study 3, I compared the effectiveness of the most and least effective fact-based reports used in Study 2 to the effectiveness of a story about a person who was wrongfully convicted due to an erroneous bitemark identification. I also examined whether a report that combines both a narrative and facts would be more effective in changing perceptions of the evidence and subsequent trial judgments than either type of information presented on its own. As in Study 2, after reading their assigned articles, participants rated the validity of a number of different types of evidence and then read a trial transcript containing either bitemark or toolmark evidence.
CHAPTER 12: STUDY TWO METHOD

Participants

Participants were 422 adults recruited on MTurk. Participants were told that they would be participating in two studies, one looking at perceptions of criminal justice in the media and one looking at juror decision-making in homicide trials (see Procedure). As such, they were compensated $2 in exchange for their participation. Seven participants who completed either the first or the second part of the study multiple times were excluded. Additionally, I excluded 122 participants who answered three or more attention/manipulation checks incorrectly in the first part of the study and/or any of the attention/manipulation checks in the second part. Thus, my final sample consisted of 293 participants. The age of participants ranged from 19 to 73 years ($M = 37.61, SD = 12.56$). There were 138 male participants (47.1%) and 144 female participants (49.1%); 11 did not provide a gender. Most of the participants were White (78.2%), but the sample also included Hispanic (5.1%), Black (7.8%), and Asian participants (4.8%). The vast majority of participants (87.6%) had completed at least some college; almost a third (30.4%) had a Bachelor’s degree and 18.4% had completed at least some graduate school. All participants were located in the United States and all but seven were U.S. citizens (excluding these seven participants did not change the pattern of results, so they were retained).

Design

The design of the study was a 5 (target article type: none, complex, simple, affirming, corrective) x 2 (trial evidence type: bitemark, toolmark) between-subjects factorial design.
Materials

Articles.

Control articles. I downloaded three articles on a range of criminal justice topics but unrelated to forensic science, jurors, or homicide cases, from LexisNexis to serve as control articles. I edited them for length only so that they would be comparable to the target articles. The articles discussed: 1) the effectiveness of mandatory minimums in reducing drug crime rates; 2) attempts by the United States to prevent human trafficking; and 3) the results of a large-scale study on sexual abuse in the Catholic Church. Participants in the control groups read all three articles, while participants in the experimental groups read two of the three (the mandatory minimums and human trafficking articles).

Target articles. I created four target articles that reported on the findings of the NAS Report (see Appendices D-G). The complex version discussed in detail the findings of the report regarding bitemark analysis and its questionable validity using the language used in the report. After giving a very basic overview of the general findings of the NAS Report regarding the forensic sciences, the article targeted bitemark evidence specifically, pointing to the dearth of research, lack of population studies, and frequently diverging conclusions of experts. Additionally, I copied the three specific problems with bitemark analysis and interpretation identified by the NAS Report regarding uniqueness, the possibility of distortion, and lack of standards for determining the value of a match (see Chapter 2) into the article. The complex report had a Flesch Reading Ease rating of 18.9 and a Flesch-Kincaid Grade Level of 17.3. The simple version opened with the same overview of the general findings of the NAS Report but I simplified the language used. When talking about bitemark analysis, it again used simpler language and targeted specifically the issue of uniqueness, discussing in detail the problem that it
is not known whether people’s dentitions are unique (including the fact that there is a lack of research on the issue). The simple report had a Flesch Reading Ease rating of 44.7 and a Flesch-Kincaid Grade Level of 13.1.

The next two versions added to the simple report to either change the framing of the issue (affirming) or to provide additional information to be used by prospective jurors (corrective). The affirming version included additional framing sections so that the focus was on the importance of the criminal justice system and how improving science would help to ensure that guilty parties are found and apprehended rather than simply detailing the problems with the science. In addition, problems with validity were framed as hampering the ability of prosecutors to convict guilty parties. In addition to the information from the simple report, the corrective version gave specific information about how bitemark evidence should be evaluated in lieu of accepting it uncritically. Of particular importance, it instructed readers about how to interpret expert testimony about a bitemark exclusion or match. Specifically, it informed them that if an expert testified that a bitemark did not match, jurors could be fairly certain that the suspect did not make the bitemark. By contrast, it informed them that if an expert testified that a bitemark did match, jurors should not conclude that the suspect made the bitemark but that “he or she is one of an unknown number of people whose teeth match certain characteristics…” Further, it explicitly informed the reader that, “a bitemark match in itself should not be considered sufficient evidence to support a verdict of guilt.”

Pilot testing. I conducted three rounds of pilot testing on the articles, one with undergraduates and two with MTurk participants.

Round 1. I conducted the first round of pilot testing with 160 undergraduates enrolled in a 200-level psychology class who participated in exchange for extra credit. Each participant read
one article, either a control or a target article. To check that they had paid attention, they then answered three multiple questions about the information they read, and made one 1-100 continuous rating of something related to the message of the article. In the case of the forensic science questions, they were asked to rate the reliability of bitemark evidence. They also attempted to answer the questions about the articles they had not read in order to ensure that the questions could not be answered correctly simply by guessing, and so that I had control ratings for the bitemark reliability question. In examining the results of the pilot testing, accuracy was so low that it was not clear that students had in fact read the articles. Only 10.4% of participants reading the target articles answered all three questions correctly, and a full 13.5% answered none of the three correctly. For the control articles, accuracy was similarly low with the percentage of participants answering all three questions correctly ranging from 9.5% to 28.6%. Accordingly, I discarded the results of the first round of pilot testing.

Round 2. I conducted a second round of pilot testing using 123 participants from MTurk who each read and answered questions about either one control and one target article or two control articles. Participants answered questions only for the articles they read about with the exception of the continuous question about the reliability of bitemark evidence, which was answered by all participants. Participants were paid $0.20 in exchange for their participation. For this round, I shortened the articles by roughly a paragraph each in hopes of increasing attention to individual articles. Additionally, I revised the wording of the questions slightly in an attempt to increase clarity, and added a manipulation check question specific to each of the target articles. Accuracy for the questions related to the target articles increased considerably: 69.4-81.5% of participants answered each of the individual forensic science questions correctly. However, the range suggested that even if most participants were attending to the articles, the
questions may not have been equally clear. Thus, I slightly revised the wording of the questions and conducted one more round of pilot testing with the goal of achieving at least 75% accuracy on all questions. I did not revise the articles further, since an initial analysis indicated that they were having their intended effect. When I restricted the pilot sample to control participants and participants who had answered at least three of the target article questions correctly, an ANOVA indicated that the articles did significantly influence ratings of the reliability of bitemark evidence, $F(4, 92) = 2.61, p = .04, \eta^2_p = 0.10$.

**Round 3.** I conducted the last round of pilot testing with 115 MTurk participants who each read one target or one control article. For the first three questions—that were relevant to all target articles—accuracy ranged from 77.7-85.4%. For the questions specific to the individual target articles accuracy was somewhat lower, ranging from 67.9-82.6%. However, mindful of the fact that these were online participants and some may simply not have fully attended to the materials, I decided to move forward with the study. I did, however, slightly revise three of the questions in a final attempt to improve clarity (see Appendix H for the final set of questions).

**Beliefs about the validity of forensic sciences questionnaire.** Participants answered three questions about a variety of different types of forensic science. They also answered similar questions about other types of evidence so as not to arouse suspicion. I did not use the more vague evidence types that were included in Study 1 (victim evidence, police evidence, and expert evidence), since I could not be sure that all participants were interpreting them the same way. I also increased the number of specific types of evidence, including two additional types of forensic science (toxicology, a laboratory-based science, and arson analysis); in total, participants were asked about nine types of evidence (see Table 11). For each evidence type, participants answered two questions to gauge perceptions of validity, one asking them to rate the
reliability of the evidence, and one asking them to rate the extent to which the evidence is based on good scientific principles, both on scales of 1 (not at all) to 100 (very). I then averaged their responses to these questions to form a composite “validity-belief” variable for each type of evidence. Participants were also asked one question about how much they knew about the type of evidence so that I was able to get a sense of the strength of their beliefs about reliability and on what they were based. This question was rated on a scale of 1 (almost nothing) to 100 (extremely knowledgeable). The order in which the different types of evidence were listed was randomized.

**Transcripts.** I used transcripts from Study 1 for Study 2. Half of the participants read the transcript containing the bitemark expert and half read the transcript containing the toolmark expert. As indicated earlier, I selected toolmark evidence as the comparison because its reliability ratings and DV measures in Study 1 were the most similar to the ratings for bitemark evidence. For both types of evidence, I used transcripts from the condition where the evidence matched the defendant and the non-substantive cross-examination was used.

**Need for Cognition Scale.** I again used the short version of the Need for Cognition Scale (Cacioppo et al., 1984). Once NC scores were calculated, I again divided the scores into thirds based on percentiles to create a low NC, moderate NC, and high NC group. Mean scores were 51.61 (SD = 10.16) for the low NC group, 70.79 (SD = 2.85) for the moderate NC group, and 81.44 (SD = 4.02) for the high NC group.

**Manipulation checks.** I included four questions regarding the articles and five questions regarding the trial materials as attention/manipulation checks. Each article was followed by four multiple choice questions about the content of the article; the four questions about the target articles served as attention/manipulation checks for part one of the study. The first three questions were relevant to all versions and gauged general attention to content, while the last
question was specific to the version of the article that had been read to ensure that participants had attended to the experimental manipulation. The five questions retained as attention/manipulation checks in Study 1 were used to check attention to the trial materials in Study 2.

**Dependent Measures**

I used the composite *validity-belief* variables for each evidence type as my dependent measures regarding pretrial beliefs. For perceptions of guilt, and perceptions of the evidence and expert, I used the same measures used in Study 1 except that verdict confidence, expert credibility, and understanding of the expert and evidence were measured on a 9-point scale rather than a 7-point scale to allow more variability in responses.

**Procedure**

Participants signed up on the MTurk site. On the site, participants read that two researchers at John Jay College of Criminal Justice were seeking participants for their research and that in order to receive compensation, they needed to participate in both studies. In order to enhance the believability of the cover story regarding two separate researchers, I presented the combining of the studies as a cost-saving measure undertaken by the researchers so that both were able to collect data from a sufficient number of participants, but at a higher rate of compensation than they would be able to provide otherwise. Participants were informed that the purpose of the first study was to examine what makes criminal justice articles interesting and how to make them more compelling, while the purpose of the second study was to examine how jurors respond to different types of testimony in criminal trials. Participants who wished to continue were directed to the first study website (www.psychsurveys.org), which contained the articles. The study site randomly assigned participants to condition. After reading informed
consent (see Appendix C), participants read their assigned articles and—following each article—responded to a series of questions about the information contained in the articles. Responses to these questions also served as attention/manipulation checks to ensure that they had attended to the articles (see Materials). They also rated how easy each article was to understand and how clearly the information in each article was reported in order to bolster my cover story.

Participants in experimental groups always read the target article second (in order to avoid primacy or recency effects). At the end of the first part of the study, participants were asked to enter their MTurk ID (this allowed me to match their first session data with their second session data). Participants were then directed to the second study website (www.qualtrics.com) where they again entered their MTurk ID, confirmed their consent to participate in the study, and responded to the pretrial evidence validity questions. They then read their assigned transcript and responded to the trial-related attention/manipulation check questions, and the post-trial perceptions of guilt and perceptions of the evidence and expert questions. Once they had completed the post-trial questions, they completed the Need for Cognition Scale and provided basic demographic information. When they had completed the study, they were given a unique participant ID number to enter into the MTurk site. I used that ID to check whether they had completed the study. Once I was able to verify that their participant ID number was valid, I released their funds.

Hypotheses

H9: Beliefs about the validity of bitemark evidence will be lower in experimental conditions than control conditions.
H10: Perceptions of guilt and of the evidence and expert will be lower in experimental conditions than control conditions. However, these differences will be greater for participants reading bitemark transcripts than toolmark transcripts.

H11: The complex article will have the least influence on perceptions and judgments. Overall, beliefs about the validity of the evidence, perceptions of guilt, and perceptions of the expert and evidence will be higher in complex article conditions than in simple, affirming, or corrective article conditions.

H12: Results will differ for high and low NC participants. For low NC participants, beliefs about the validity of the evidence, perceptions of guilt, and perceptions of the expert and evidence will be higher in complex article conditions than in simple, affirming, or corrective article conditions, matching the overall pattern. For high NC participants, there will be no difference between article types.
CHAPTER 13: STUDY TWO RESULTS

In analyzing the Study 2 data, I again organized my results based on my hypotheses. For verdicts, I used logistic regression; for all other analyses, I used ANOVA or MANOVA using Tukey tests for all post hoc follow-ups. With the exception of initial analyses involving validity-beliefs, all models included article type and evidence type unless otherwise specified (participants had not read the transcripts when they made their validity-belief ratings, so evidence type was not relevant).

Hypothesis 9: Beliefs about the validity of bitemark evidence will be lower in experimental conditions than control conditions.

My ninth hypothesis was supported: all four articles significantly reduced perceptions of the validity of bitemark evidence. Further, the articles had some influence on perceptions of other types of evidence.

I first examined pretrial validity-belief ratings across all evidence types in the control group. In the control group, the forensic sciences were rated more highly than other types of evidence (see Table 11). DNA was rated as most valid, followed by toxicology, fingerprint, bitemark, arson, toolmark, polygraph, confession, and finally eyewitness evidence, which was rated as least valid. There was a great deal of variability in ratings, however, with high standard deviations for all evidence types.

I next examined the validity-belief ratings for bitemark evidence as a function of article type. An ANOVA revealed that article type did influence validity ratings, \( F(4, 288) = 22.48, p < .001, \eta^2_p = .24 \). Post hoc Tukey tests indicated that bitemark validity-belief ratings for participants in the control group were significantly higher than for those who read the complex
(p < .001, $M_{\text{diff}} = 37.75$ [95% CI: 25.71, 49.78]), simple ($p < .001, M_{\text{diff}} = 29.54$ [95% CI: 17.27, 41.82]), affirming ($p < .001, M_{\text{diff}} = 28.22$ [95% CI: 15.74, 40.69]), or corrective articles ($p < .001, M_{\text{diff}} = 28.75$ [95% CI: 16.93, 40.58]) (for means see Table 12). None of the experimental groups differed significantly (all $ps > .2$).

I also found that the target articles significantly influenced ratings of the other types of evidence examined, Wilks’ $\lambda = .81$, $F(32, 1038) = 1.94$, $p = .001$, $\eta_p^2 = .05$. Indeed, follow-up analyses indicated that article type significantly influenced perceptions of toolmark, $F(4, 288) = 6.17$, $p < .001$, $\eta_p^2 = .08$, polygraph, $F(4, 288) = 4.05$, $p = .003$, $\eta_p^2 = 0.05$, toxicology, $F(4, 288) = 2.69$, $p = .03$, $\eta_p^2 = .04$, and arson evidence, $F(4, 288) = 3.38$, $p = .01$, $\eta_p^2 = .05$. Follow-up analyses indicated that for these evidence types, participants in the control group generally gave higher ratings than participants reading at least one other type of article, but the patterns were not consistent (see Table 12). Only for bitemark evidence did all of the target articles significantly reduce validity-belief ratings.

**Hypothesis 10:** Perceptions of guilt and of the evidence and expert will be lower in experimental conditions than control conditions. However, these differences will be greater for participants reading bitemark transcripts than toolmark transcripts.

**Hypothesis 11:** The complex article will have the least influence on perceptions and judgments. Overall, beliefs about the validity of the evidence, perceptions of guilt, and perceptions of the expert and evidence will be higher in complex article conditions than in simple, affirming, or corrective article conditions.

My tenth hypothesis was partially supported. The articles did not influence verdicts or estimates of the likelihood that the defendant murdered the victim. However, they did influence
perceptions of the likelihood that the defendant was the source of the evidence in addition to perceptions of the credibility of the expert and the quality of the evidence. That said, I only found differential effects of the articles in ratings of the helpfulness of bitemark and toolmark evidence: the articles influenced ratings of bitemark evidence helpfulness but not toolmark evidence helpfulness. My eleventh hypothesis was not supported: the complex article appeared to have the greatest, rather than the least, influence on participants’ perceptions and judgments. Instead, the simple article seemed to have the least impact.

**Perceptions of guilt.** For verdict choice, a logistic regression including article type, evidence type, and their interaction was not significant, \( \chi^2(9) = 15.55, p = .08, R^2_{\text{adj}} = .07 \). Similarly, an ANOVA indicated that neither article type, \( F(4, 283) = 0.55, p = .70, \eta^2_p = .008 \), evidence type, \( F(1, 283) = 0.13, p = .72, \eta^2_p < .001 \), nor their interaction, \( F(4, 283) = 0.77, p = .55, \eta^2_p = .01 \), were related to verdict confidence.

Estimates of the probability that the defendant murdered the victim mirrored verdicts; however, the articles did influence estimates of the probability that the defendant was the source of the evidence. Neither article type, \( F(4, 283) = 1.36, p = .25, \eta^2_p = .02 \), evidence type, \( F(1, 283) = 0.05, p = .83, \eta^2_p < .001 \), nor their interaction, \( F(4, 283) = 0.31, p = .87, \eta^2_p = .004 \), were related to estimates of the probability that the defendant murdered the victim. For estimates of the probability that the defendant was the source of the evidence, an ANOVA revealed a main effect of article type, \( F(4, 283) = 3.55, p = .008, \eta^2_p = .05 \); there was no significant effect of evidence type, \( F(1, 283) = 1.24, p = .27, \eta^2_p = .005 \), and the interaction between article and evidence type did not reach significance, \( F(4, 283) = 1.57, p = .18, \eta^2_p = .02 \). Post-hoc Tukey tests indicated that source estimates were significantly higher in the control group than they were in the affirming \((p = .02, M_{\text{diff}} = 12.92 [95\% \text{ CI: 1.64, 24.21}])\) or the corrective group \((p = .009, \)
$M_{\text{diff}} = 12.86$ [95% CI: 2.28, 23.44]), and marginally higher than in the complex group ($p = .06, M_{\text{diff}} = 10.67$ [95% CI: -0.26, 21.60]); they did not differ from the simple group ($p = .30, M_{\text{diff}} = 7.87$ [95% CI: -3.29, 19.02]) (for means see Table 13). None of the experimental groups differed significantly (all $ps > .7$).

**Perceptions of expert credibility.** There was a significant main effect of evidence type, $F(4, 283) = 4.50, p = .04, \eta^2_p = .02$, and a marginally significant main effect of article type, $F(1, 283) = 2.41, p = .05, \eta^2_p = .03$, on participants’ ratings of expert credibility. The interaction was not significant, $F(4, 283) = 1.09, p = .36, \eta^2_p = .02$. The expert was rated as significantly more credible in the control group than he was in the complex group ($p = .03, M_{\text{diff}} = 0.66$ [95% CI: 0.04, 1.27]) (for means see Table 13); no other differences between article types reached significance (all $ps > .2$). The toolmark expert ($M = 7.49, SD = 1.18$) was rated as more credible than the bitemark expert ($M = 7.16, SD = 1.34$).

**Perceptions of evidence quality.** Both article type and evidence type influenced perceptions of the quality of the evidence, and there were some differential effects of article type for the two different types of evidence. There were significant effects of article type, Wilks’ $\lambda = .78$, $F(8, 564) = 9.28, p < .001, \eta^2_p = .12$, and evidence type, Wilks’ $\lambda = .95, F(2, 282) = 7.15, p = .001, \eta^2_p = .05$, on perceptions of evidence quality. Article type significantly influenced ratings of evidence validity, $F(4, 283) = 18.41, p < .001, \eta^2_p = .21$, and evidence helpfulness, $F(4, 283) = 2.54, p = .04, \eta^2_p = .04$. Ratings of evidence validity were significantly higher in the control group than they were in the complex ($p < .001, M_{\text{diff}} = 1.22$ [95% CI: 0.78, 1.68]), simple ($p < .001, M_{\text{diff}} = 0.99$ [95% CI: 0.51, 1.46]), affirming ($p < .001, M_{\text{diff}} = 1.08$ [95% CI: 0.60, 1.56]), or corrective groups ($p < .001, M_{\text{diff}} = 1.17$ [95% CI: 0.71, 1.62]) (for means see Table 13). None of the experimental groups differed significantly (all $ps > .6$). Ratings of expert helpfulness
tended to be higher in the control group than they were in the complex \((p = .08, M_{\text{diff}} = 0.58 \ [95\% \ CI: -0.04, 1.20])\) or the corrective groups \((p = .07, M_{\text{diff}} = 0.58 \ [95\% \ CI: -0.03, 1.19])\), but the differences did not reach significance; they did not differ from the simple \((p = .52, M_{\text{diff}} = 0.36 \ [95\% \ CI: -0.27, 0.99])\) or affirming groups \((p = .92, M_{\text{diff}} = 0.19 \ [95\% \ CI: -0.44, 0.84])\), and none of the experimental groups differed significantly \((all \ p > .4)\).

Evidence type significantly influenced ratings of evidence validity, \(F(1, 283) = 13.61, p < .001, \eta^2_p = .05\), but not helpfulness, \(F(1, 283) = 0.68, p = .41, \eta^2_p = .002\). Toolmark evidence \((M = 4.29, SD = 1.07)\) was rated as significantly more valid than bitemark evidence \((M = 3.85, SD = 1.14)\). The interaction between article type and evidence type did not reach significance, Wilks’ \(\lambda = .95, F(8, 564) = 1.71, p = .09, \eta^2_p = .02\); however, since I was expecting an interaction, I examined the univariate tests as well. I found that there was a significant interaction between article type and evidence type on ratings of expert helpfulness, \(F(4, 283) = 2.44, p = .047, \eta^2_p = .03\), but the interaction did not reach significance for ratings of expert validity, \(F(4, 283) = 2.10, p = .08, \eta^2_p = .03\). For ratings of evidence helpfulness, I found that there was a main effect of article type on ratings of the helpfulness of bitemark, \(F(4, 153) = 3.51, p = .009, \eta^2_p = .08\), but not toolmark evidence, \(F(4, 130) = 1.48, p = .21, \eta^2_p = .04\). Bitemark evidence was rated as significantly more helpful in the control condition \((M = 4.26, SD = 1.18)\) than it was in the complex condition \((M = 3.16, SD = 1.29; p = .004, M_{\text{diff}} = 1.10 \ [95\% \ CI: 0.25, 1.95])\). There was no difference between the control condition and the simple \((p = .89, M_{\text{diff}} = 0.27 \ [95\% \ CI: -0.55, 1.10])\), affirming \((M = 3.79; SD = 1.26; p = .58, M_{\text{diff}} = 0.47 \ [95\% \ CI: -0.41, 1.34])\), or corrective conditions \((M = 3.67, SD = 1.16; p = .92, M_{\text{diff}} = 0.59 \ [95\% \ CI: -0.21, 1.39])\), and no differences between experimental groups reached significance \((all \ p > .07)\).
Perceived understanding of the expert and evidence. The articles did not influence perceived understanding of the expert and evidence. There was not a main effect of article type, $F(4, 283) = 1.14, p = .34, \eta_p^2 = .02$, or evidence type, $F(1, 283) = 0.03, p = .87, \eta_p^2 < .001$, nor was there a significant interaction, $F(4, 283) = 0.99, p = .41, \eta_p^2 = .01$, on participants’ ratings of how well they understood the expert’s testimony. Similarly, there was not a main effect of article type, $F(4, 283) = 0.98, p = .42, \eta_p^2 = .02$, or evidence type, $F(1, 283) = 0.09, p = .76, \eta_p^2 < .001$, nor was there a significant interaction, $F(4, 283) = 0.70, p = .59, \eta_p^2 = .01$, on participants’ ratings of how well they understood the forensic technique used by the expert.

Given that some differences between article types were found for other DVs, I also checked to see whether participants’ pretrial ratings of how easy or difficult it was for them to understand the articles differed dependent on article type. However, an ANOVA did not indicate any differences in perceived understanding of the articles based on target article type, $F(3, 221) = 0.77, p = .51, \eta_p^2 = .01$. Similarly, participants’ ratings of the clarity of the information presented in the articles did not differ based on target article type, $F(3, 221) = 0.50, p = .69, \eta_p^2 = .007$.

Hypothesis 12: Results will differ for high and low NC participants. For low NC participants, beliefs about the validity of the evidence, perceptions of guilt, and perceptions of the expert and evidence will be higher in complex article conditions than in simple, affirming, or corrective article conditions, matching the overall pattern. For high NC participants, there will be no difference between article types.

My hypothesis regarding NC was not supported. Only one interaction between NC and article type was found: for source likelihood estimates, low NC participants were influenced by
the affirming article, while high NC participants were influenced by the corrective article.

Indeed, NC had very little influence on any dependent measures: I found a direct influence only on perceived understanding of the expert. For analyses of NC, I only report effects or interactions not involving NC when they are different from in the primary analyses. Otherwise, only main effects or interactions with NC are reported. However, with the exception of the analysis of validity-beliefs, all models included NC group, article type, and evidence type.

**Pretrial validity beliefs.** An ANOVA revealed that there was no effect of NC on perceptions of the validity of bitemark evidence, $F(2, 268) = 1.72, p = .18, \eta_p^2 = .01$, nor was there a significant interaction between NC and article type, $F(8, 268) = 0.88, p = .53, \eta_p^2 = .03$. Similarly, there was no effect of NC on perceptions of the validity of other types of evidence, Wilks’ $\lambda = .94, F(16, 522) = 1.05, p = .40, \eta_p^2 = .03$, and no interaction between NC and article type, Wilks’ $\lambda = .77, F(64, 1512) = 1.12, p = .25, \eta_p^2 = .03$.

**Perceptions of guilt.** For verdicts, a logistic regression including NC, article type, evidence type, and all interactions with NC, was not significant, $\chi^2(25) = 26.53, p = .38$, $R^2_{nag} = .12$. For verdict confidence, there was not a significant main effect of NC, $F(2, 253) = 2.60, p = .08, \eta_p^2 = .02$, and no interactions with NC reached significance (all $p$s > .1).

There was no effect of NC on estimates of the probability that the defendant murdered the victim, $F(2, 253) = 0.33, p = .72, \eta_p^2 = .003$, and no interactions reached significance (all $p$s > .3). There was also no main effect of NC on estimates of the probability that the defendant was the source of the evidence, $F(2, 253) = 0.13, p = .88, \eta_p^2 = .001$, but there was a significant interaction between NC and article type, $F(2, 253) = 2.01, p = .045, \eta_p^2 = .06$. Follow-up analyses indicated that there was a significant effect of article type on source probability estimates for low NC, $F(4, 87) = 3.51, p = .01, \eta_p^2 = .14$, and high NC participants, $F(4, 90) =$
3.08, \( p = .02, \eta^2_p = .12 \), but there was not for moderate NC participants, \( F(4, 91) = 1.42, p = .23, \eta^2_p = .06 \). Further analyses indicated that for low NC participants, source estimates were significantly higher in the control group than they were in the affirming group (\( p = .02, M_{\text{diff}} = 20.48 \) [95% CI: 1.86, 39.11]), while for high NC participants, source estimates were significantly higher in the control group than there were in the corrective group (\( p = .01, M_{\text{diff}} = 20.17 \) [95% CI: 3.11, 37.22]) (for means see Table 14). No other differences reached significance (all \( ps > .1 \)).

**Perceptions of the expert and evidence.** There was not a main effect of NC on expert credibility, \( F(2, 253) = 1.66, p = .19, \eta^2_p = .01 \), and no interactions with NC reached significance (all \( ps > .065 \)). Additionally, the main effect of evidence type was no longer significant, \( F(1, 253) = 1.96, p = .16, \eta^2_p = .008 \), while the main effect of article type changed from marginal to significant, \( F(4, 253) = 2.64, p = .04, \eta^2_p = .04 \) (for follow-ups, see analyses above). I did not find a significant effect of NC on perceptions of evidence quality, Wilks’ \( \lambda = .98, F(4, 504) = 1.44, p = .22, \eta^2_p = .01 \), and no interactions with NC reached significance (all \( ps > .1 \)).

There was a main effect of NC on participants’ ratings of how well they understood the expert’s testimony, \( F(2, 253) = 6.66, p = .002, \eta^2_p = .05 \), but no interactions with NC reached significance (all \( ps > .3 \)). High NC participants gave significantly higher ratings to their understanding of the expert’s testimony than low NC (\( p < .001, M_{\text{diff}} = 0.77 \) [95% CI: 0.26, 1.28]) or moderate NC participants (\( p = .01, M_{\text{diff}} = 0.61 \) [95% CI: 0.11, 1.12]) (for means see Table 15). There was a marginally significant effect of NC on participants’ ratings of how well they understood the forensic technique used by the expert, \( F(2, 253) = 2.85, p = .06, \eta^2_p = .02 \), but no interactions with NC reached significance (all \( ps > .065 \)). Follow-ups indicated that high NC participants tended to give higher ratings to their understanding of his forensic technique.
than low NC participants, but the difference did not reach significance ($p = .08$, $M_{\text{diff}} = 0.59$ [95% CI: -0.05, 1.23]); neither significantly differed from moderate NC participants’ ratings (all $ps > .1$). 
CHAPTER 14: STUDY TWO DISCUSSION

All four of the fact-based articles targeting bitemark evidence were successful in reducing perceptions of the validity of bitemark evidence. Compared to the control group, the complex, simple, affirming, and corrective articles all significantly reduced validity-belief ratings, the composite measure assessing perceptions of reliability and scientific quality. There was also some effect of the target articles on perceptions of several other evidence types: toolmark, toxicology, arson, and polygraph evidence, the first three of which are forensic disciplines. This result suggests that reading the articles may have had some influence on perceptions of forensic science more broadly, not only perceptions of the evidence specifically targeted in the article. No other type of forensic science (other than DNA) was mentioned by name, and specific criticisms were given for bitemark evidence only. Thus it is possible that inducing skepticism about one type of evidence tends to make people more skeptical regarding other types of evidence. However, the articles did mention that the NAS Report had found problems with all non-DNA disciplines, which also may account for the reduced perceptions of some of the other forensic disciplines. That said, different articles influenced ratings of the different evidence types and there were no consistent patterns, so at this point it is not clear to what extent reading fact-based articles influences perceptions of forensic science in general.

Consistent with previous research (Hans et al., 2011; Lieberman et al., 2008) and with the results of Study 1, participants who had not read any of the target articles rated forensic science evidence more highly than other types of evidence. Thus, these results add additional support to the argument that people may tend to have inflated perceptions of forensic science evidence and that education regarding its deficiencies is needed. However, again consistent with prior research and Study 1, these results suggest that people do have some knowledge of the comparative
validity of different types of forensic science, although certain types may be given too much
dereference. Interestingly, participants gave quite high validity-belief ratings to toxicology, the
non-DNA laboratory-based discipline, suggesting that people may on some level be aware that
laboratory disciplines have greater scientific value. However, fingerprint evidence was given
comparable ratings to toxicology, giving further support to the notion that its scientific validity
tends to be overestimated.

Consistent with the literature on PTP (Spano et al., 2011; Stebley et al., 1999), reading
the articles prior to trial materials influenced participants’ perceptions of trial evidence. While
the target articles did not consistently reduce perceptions of guilt, they did reduce perceptions of
the expert and the evidence. There was no effect of the articles on verdict, verdict confidence, or
estimates of the likelihood that the defendant murdered the victim. The articles did, however,
have some influence on estimates of the likelihood that the defendant was the source of the
evidence. Further, they significantly influenced perceptions of the credibility of the expert and
the quality of the evidence. Given that the articles reported facts about forensic science only and
did not give specific information about the case, these results add to the literature showing that
pretrial media need not be emotional in nature (Kramer & Kerr, 1989; Wilson & Bornstein,
1998) or specific to the case to be litigated (Greene & Loftus, 1084; Greene & Wade, 1988) in
order to influence jurors. In fact, the effect was somewhat broader than I had expected: even
though they targeted bitemark evidence, the articles did not appear to influence perceptions of
bitemark evidence to a greater extent than those of toolmark evidence. Indeed, when the articles
had an effect, they almost invariably influenced perceptions for participants in both the bitemark
and toolmark conditions. Only once did I find a significant interaction between article type and
evidence type: the articles had a greater influence on perceptions of the helpfulness of bitemark
than toolmark evidence. It is possible that this failure to find a difference was at least partially due to the articles having some effect on perceptions of forensic science more broadly. Indeed, as stated above, the articles did have some influence on perceptions of several other types of evidence, and both the complex and simples articles significantly reduced perceptions of the validity of toolmark evidence. It is worth noting, however, that toolmark evidence was generally viewed more positively than bitemark evidence in Study 2, which stands in direct contrast to Study 1, where there were no differences between the two in any post-trial ratings. This discrepancy suggests that the articles as a whole may have, at least to some extent, influenced perceptions of bitemark evidence more than toolmark evidence.

Contrary to my expectations and to research showing that simplicity is preferred to complexity (Lewandowsky et al., 2012; Lombrozo, 2007), the complex article appeared to have the greatest influence on participants, while the simple article had the least influence of the four articles examined. The complex article was the only target article that influenced perceptions of expert credibility and the helpfulness of bitemark evidence. Further, the only differences I found between experimental groups were between the complex and simple conditions, and in each case, the complex article had a greater influence than the simple. In fact, the simple article was the only article that did not even marginally influence source estimates, and appeared to have the least influence on ratings of the quality of bitemark evidence. The complex article was the only target article that attacked bitemark evidence from multiple angles, which may help to account for its greater influence, particularly if it was not perceived as overly complex or difficult to understand. Participants who read the complex article did not, in fact, rate it as more difficult to understand or as less clear than participants who read the other articles, despite the fact that readability statistics were quite different for the complex and simple articles (see Method).
While the affirming article did appear to have a greater influence than the simple article, it did not appear to have a greater influence than the complex article. The affirming article was included in light of previous research showing the success of corrections to misinformation that in some way affirm people’s previously held beliefs or that are framed to be consistent with their worldviews (Hardisty et al., 2010; Lewandowsky et al., 2012). Thus, the affirming article framed the findings of the NAS Report in such a way as to attempt to affirm people’s beliefs in the importance of the criminal justice system and the need to convict guilty parties. The affirming article did influence perceptions of the likelihood that the defendant was the source of the evidence and perceptions of the validity of the evidence. However, there was no evidence that it did so more than other articles. Further, it did not influence perceptions of expert credibility or evidence helpfulness, which were influenced by other articles. It is possible that in part it did not have a greater influence because people’s views about bitemark evidence are not particularly strongly held. Indeed, control participants’ pretrial ratings of their knowledge of bitemark evidence suggest that people do not feel that they are particularly knowledgeable about bitemark evidence. Thus, it seems likely that attacking the validity of bitemark evidence is not particularly threatening to people’s beliefs.

Given that the corrective article provided specific information about how to evaluate bitemark evidence, it is surprising that it did not have a greater influence on participants’ perceptions of the trial evidence. The corrective article specifically instructed participants that an expert’s testimony that the defendant was the source of the evidence did not mean that he was the source of the evidence. Thus, it is particularly surprising that the corrective article did not influence estimations of the likelihood that the defendant was the source of the evidence to a greater extent than the affirming or complex articles did. Research on correcting misinformation
suggests that providing an alternative to misinformation is more effective than simply correcting it (H. M. Johnson & Seifert, 1994; Seifert, 2002). With that in mind, the purpose of the corrective article was to provide participants with information about evaluating evidence that they could use in lieu of simply accepting expert testimony. However, especially given their low knowledge ratings, it is possible that participants may not have had beliefs about how bitemark evidence should be evaluated prior to readings the articles. Further, it is possible that participants needed a stronger or different type of “alternate” (e.g., an alternate suspect; H. M. Johnson & Seifert, 1994) for this method to be particularly effective.

There is some evidence that the corrective article may have been more effective for high NC participants, given that it influenced only high NC participants’ estimates of the likelihood that the defendant was the source of the evidence. By contrast, the affirming article was what influenced low NC participants’ source likelihood estimates. It is possible that high NC participants’ greater motivation to systematically process information made them more motivated to use the information provided to them in the corrective article. By contrast, perhaps low NC participants are more likely to rely on gut feelings rather than careful evaluation of the evidence. Further, the results of Study 1 suggest that people low in NC may be more punitive, while having more reservations about science. If true, it is possible that appealing to low NC participants’ feelings about criminal justice, rather than just providing factual scientific information, was more effective than it was for high NC participants. That said, given that I did not find differences between articles types for different NC groups on any other measures, it is not clear to what extent either might be true. Indeed, generally speaking, there were almost no differences based on NC in Study 2, suggesting that low, moderate, and NC participants generally responded similarly. As in Study 1, high NC participants reported that they understood
the expert’s testimony better than low NC participants. However, with the exception of source probability estimates, I found no differences between NC groups in their perceptions of guilt or in perceptions of the quality of the evidence or the credibility of the expert.

Taken as a whole, the results of Study 2 suggest that factual articles can effectively influence perceptions of the validity of bitemark evidence. Indeed, they influenced both pretrial ratings of the validity of bitemark evidence in general, and post-trial ratings of the quality of bitemark evidence presented in a criminal trial. However, they were not able to influence participants’ ultimate judgments or how likely they felt it was that the defendant had committed the crime. Thus, in Study 3, I investigated the possibility that a story, or a story supplemented by facts, might be able to influence not only perceptions of the evidence, but also perceptions of guilt.
CHAPTER 15: STUDY THREE METHOD

Participants

Participants were 428 adults recruited on MTurk who, as in Study 2, were told that they would participate in two different studies and compensated $2 for their participation. Eleven participants completed either the first or partially completed the second part of the study multiple times and were therefore excluded. Additionally, I excluded 126 participants who answered three or more forensic science or any story manipulation checks in the first part of the study and/or any of attention/manipulation checks incorrectly in the second part. Thus my final sample consisted of 291 participants. The age of participants ranged from 19 to 75 years (\(M = 36.50, SD = 11.99\)). There were 123 male (42.3%) and 165 female participants (56.7%); three participants did not indicate their gender. Most of the participants were White (79.0%), but the sample also included Hispanic (4.8%), Black (5.8%), and Asian participants (4.5%). The vast majority of participants (92.0%) had completed at least some college; almost a third (32.6%) had a Bachelor’s degree and 20.6% had completed at least some graduate school. All participants were located in the United States and all but eight were U.S. citizens (excluding these participants did not change the pattern of results, so I retained them).

Design

The design of the study was a 5 (target article type: none, complex, simple, story, story-and-fact) x 2 (trial evidence type: bitemark, toolmark) between-subjects factorial design.

Materials

Articles.

Control articles. I used the same control articles in Study 3 that I had used in Study 2.
Target articles. The first two target articles were fact-based. I used the complex article and the simple article since they were the most and least effective, respectively, in Study 2. The second two articles contained a narrative (see Appendices I-J). The story article talked about the Robert Lee Stinson case described in Chapter 1. It described his conviction, focusing on the role of the bitemark testimony, and then his exoneration due to DNA evidence. I selected his case because the bitemark evidence was the only evidence connecting him to the crime, and because the DNA evidence specifically contradicted the findings of the bitemark experts: Stinson was excluded as a possible contributor to the saliva from the bitemarks. The story-and-fact article first discussed the Stinson case and then supplemented the story with the information from the simple report that specifically targeted problems with bitemark evidence (i.e., focused on the problem of uniqueness using simple language).

Pilot testing. I pilot tested the story articles with 75 participants recruited via MTurk. They were compensated $0.20 for their participation. Each participant read either the story article or one of two possible versions of the story-and-fact article: one that supplemented the story with information from the complex version of the fact-based article from Study 2, and one that supplemented it with information from the simple version. Participants then answered four multiple choice and one continuous 1-100 scale question regarding the reliability of bitemark evidence. Of the multiple choice questions, the first three asked about information contained in the story in all three conditions. For the story article, I asked one additional question about the Stinson case. For the version of the story article that supplemented the story with information from the simple report, I used the manipulation check question for the simple article. For the version of the story article that supplemented the story with information from the complex report, I used the manipulation check question for the complex article. Accuracy for the four
questions ranged from 64.3% to 96.4%. Accuracy was above 75% for three questions, including the manipulation check question for each report. When I read over the question that participants were most likely to get wrong, two response options both seemed plausible thus, I simply replaced that question with another question about the story in general and moved forward with data collection (for the final set of questions see Appendix K). I found no differences in reliability ratings for participants who read the two possible versions of the story-and-fact article (that supplemented the story with facts from the complex versus facts from the simple article). Thus, I used the version which included the information from the simple article since it was focused on just one issue.

**Beliefs about the validity of forensic sciences questionnaire.** I used the same questions and evidence types as in Study 2.

**Transcripts.** I used the same transcripts for Study 3 that I had used in Study 2—transcripts from the evidence-match, non-substantive cross-examination conditions.

**Need for Cognition Scale.** I again used the short version of the Need for Cognition Scale (Cacioppo et al., 1984), and divided the resulting scores into thirds to create low, moderate, and high NC groups. Mean scores were 50.41 (SD = 11.38) for the low NC group, 68.75 (SD = 2.43) for the moderate NC group, and 80.43 (SD = 5.06) for the high NC group.

**Manipulation checks.** As in Study 2, four multiple choice questions about the target articles served as attention/manipulation checks for the articles. For the fact-based articles, I used the same questions I used in Study 2. For the story and story-and-fact articles, the first three questions were the same and asked about information from the Stinson case. For the fourth question, I asked one additional question about the Stinson case for the story article, and then used the manipulation check question from the simple article for the story-and-fact article. As in
Study 2, I retained the five questions used as attention/manipulation checks in Study 1 to check attention to the trial materials in Study 3.

**Dependent Measures**

I used the same measures in Study 3 that I used in Study 2.

**Procedure**

I used the same procedure in Study 3 that I used in Study 2.

**Hypotheses**

H13: Overall, the story-and-fact article will have the greatest influence, but the story article will have a greater influence than either fact-based article. Within the experimental conditions, beliefs about the validity of the evidence, perceptions of guilt, and perceptions of the expert and evidence will be highest in simple article conditions and lowest in story-and-fact article conditions.

H14: There will be an interaction between article type and trial evidence type such that the story and story-and-fact reports will have an equivalent effect on perceptions of validity, guilt and of the expert and evidence in both trial evidence conditions, while the fact-based reports will have a greater impact on participants in the bitemark than the toolmark conditions.

H15: Results will differ for high and low NC participants. For low NC participants, beliefs about the validity of the evidence, perceptions of guilt, and perceptions of the expert and evidence will be highest in simple article conditions and lowest in story-and-fact article conditions, matching the overall pattern. For high NC participants, there will be no difference between the two fact-based article conditions, but both will be significantly higher than the story and story-and-fact article conditions.
CHAPTER 16: STUDY THREE RESULTS

In analyzing the Study 3 data, I again organized my results based on my hypotheses. For verdicts, I used logistic regression; for all other analyses, I used ANOVA or MANOVA using Tukey tests for all post hoc follow-ups. As in Study 2, with the exception of my initial analyses involving validity-beliefs, all models included article type and evidence type.

Hypothesis 13: Overall, the story-and-fact article will have the greatest influence, but the story article will have a greater influence than either fact-based article. Within the experimental conditions, beliefs about the validity of the evidence, perceptions of guilt, and perceptions of the expert and evidence will be highest in simple article conditions and lowest in story-and-fact article conditions.

Hypothesis 14: There will be an interaction between article type and trial evidence type such that the story and story-and-fact reports will have an equivalent effect on perceptions of validity, guilt and of the expert and evidence in both trial evidence conditions, while the fact-based reports will have a greater impact on participants in the bitemark than the toolmark conditions.

My thirteenth hypothesis was only partially supported, and my fourteenth hypothesis was not supported. The story-and-fact article did have the greatest influence: it was the only article that influenced source likelihood estimates and it also significantly influenced perceptions of the expert and evidence (although for the latter, the complex article produced comparable ratings). Unexpectedly, the story article had the least influence. While it marginally reduced pretrial perceptions of the validity of bitemark evidence, participants who read the story article did not differ from the control group on any perceptions of guilt or perceptions of the expert or evidence.
measures. However, I did not find evidence that the story-based articles were more likely to influence perceptions of both bitemark and toolmark evidence. In fact, there was some evidence that the complex article was more likely to generalize than either story-based report.

**Pretrial validity beliefs.** I first examined pretrial validity-belief ratings across all evidence types in the control group, and found similar results to Study 2 (see Table 16). I then investigated whether my target articles influenced perceptions of bitemark evidence. An ANOVA revealed that there was a significant main effect of article type on validity-belief ratings, $F(4, 286) = 38.92$, $p < .001$, $\eta^2_p = .35$. Post hoc Tukey tests indicated that bitemark validity-belief ratings for participants in the control group were significantly higher than for those who read the complex ($p < .001$, $M_{\text{diff}} = 39.83$ [95% CI: 27.21, 52.45]), simple ($p < .001$, $M_{\text{diff}} = 38.93$ [95% CI: 26.31, 51.56]), or story-and-fact articles ($p < .001$, $M_{\text{diff}} = 41.19$ [95% CI: 29.75, 52.64]) (for means see Table 17). They were only marginally higher than for those who read the story article ($p = .05$, $M_{\text{diff}} = 11.48$ [95% CI: -0.09, 23.05]). Additionally, validity-belief ratings were significantly higher for those who read the story article than for those who read the complex ($p < .001$, $M_{\text{diff}} = 28.35$ [95% CI: 15.77, 40.93]), simple ($p < .001$, $M_{\text{diff}} = 27.46$ [95% CI: 14.88, 40.03]), or story-and-fact articles ($p < .001$, $M_{\text{diff}} = 29.72$ [95% CI: 18.32, 41.11]).

I also found that the target articles significantly influenced ratings of the other types of evidence examined, Wilks’ $\lambda = .84$, $F(32, 1030) = 1.52$, $p = .03$, $\eta^2_p = .04$. Follow-up analyses indicated that article type significantly influenced perceptions of toolmark, $F(4, 286) = 4.80$, $p = .001$, $\eta^2_p = .06$, polygraph, $F(4, 286) = 3.10$, $p = .02$, $\eta^2_p = .04$, and arson evidence, $F(4, 286) = 2.85$, $p = .02$, $\eta^2_p = .04$, and marginally influenced perceptions of fingerprint evidence, $F(4, 288) = 2.40$, $p = .05$, $\eta^2_p = .03$. Follow-up analyses did not reveal consistent patterns and in some cases revealed differences between target articles, rather than with the control group (see Table
17). Only for bitemark evidence did all of the target articles consistently reduce validity-belief ratings.

**Perceptions of guilt.** For verdict choice, a logistic regression including article type, evidence type, and their interaction was not significant, $\chi^2(9) = 5.62, p = .78, R^2_{\text{neg}} = .03$. Similarly, an ANOVA revealed that neither article type, $F(4, 281) = 1.71, p = .15, \eta_p^2 = .02$, evidence type, $F(1, 281) = 0.06, p = .81, \eta_p^2 < .001$, nor their interaction, $F(4, 281) = 1.30, p = .27, \eta_p^2 = .02$, were related to verdict confidence. Estimates of the probability that the defendant murdered the victim mirrored verdicts. Neither article type, $F(4, 281) = 0.77, p = .55, \eta_p^2 = .01$, evidence type, $F(1, 281) = 2.83, p = .09, \eta_p^2 = .01$, nor their interaction, $F(4, 281) = 0.93, p = .45, \eta_p^2 = .01$, were related to estimates of the probability that the defendant murdered the victim.

For estimates of the probability that the defendant was the source of the evidence, there was no main effect of article type, $F(4, 280) = 1.40, p = .23, \eta_p^2 = .02$, or evidence type, $F(1, 280) = 2.65, p = .11, \eta_p^2 = .009$, but there was a significant interaction between the two, $F(4, 280) = 3.49, p = .008, \eta_p^2 = .05$. Follow-up analyses showed that article type significantly influenced source estimates for bitemark evidence, $F(4, 137) = 3.70, p = .007, \eta_p^2 = .10$, but not for toolmark evidence, $F(4, 143) = 1.21, p = .31, \eta_p^2 = .03$. Post-hoc Tukey tests indicated that in bitemark conditions, estimates were significantly higher in the control group than they were in the story-and-fact group ($p = .002, M_{\text{diff}} = 23.02 [95\% CI: 6.05, 39.99]$). They did not differ significantly from the complex ($p = .18, M_{\text{diff}} = 15.11 [95\% CI: -3.64, 33.85]$), simple ($p = .29, M_{\text{diff}} = 13.20 [95\% CI: -5.30, 31.71]$), or story group ($p = .49, M_{\text{diff}} = 9.68 [95\% CI: -6.94, 26.29]$) (for means see Table 18). None of the experimental groups were significantly different (all $ps > .1$).
Perceptions of expert credibility. There were significant main effects of article type, $F(4, 281) = 4.10, p = .003, \eta_p^2 = .06$, and of evidence type, $F(1, 281) = 8.29, p = .004, \eta_p^2 = .03$, on participants’ ratings of expert credibility. The interaction was not significant, $F(4, 281) = 1.51, p = .20, \eta_p^2 = .02$. The expert was rated as significantly more credible in the control group than he was in the complex group ($p = .006, M_{\text{diff}} = 0.89 [95\% \text{ CI: 0.18, 1.59}]$) or the story-and-fact group ($p = .04, M_{\text{diff}} = 0.67 [95\% \text{ CI: 0.03, 1.30}]$) (for means see Table 19); the control group did not differ from the simple ($p = .34, M_{\text{diff}} = 0.48 [95\% \text{ CI: -0.23, 1.18}]$), or story groups, ($p = .77, M_{\text{diff}} = 0.27 [95\% \text{ CI: -0.37, 0.92}]$) and no differences between experimental groups reached significance (all $p_s > .1$). The toolmark expert was rated as more credible ($M = 7.38, SD = 1.24$) than the bitemark expert ($M = 6.96, SD = 1.49$).

Perceptions of evidence quality. There were significant main effects of article type, Wilks’ $\lambda = .86, F(8, 560) = 5.58, p < .001, \eta_p^2 = .07$, and evidence type, Wilks’ $\lambda = .89, F(2, 280) = 17.26, p < .001, \eta_p^2 = .11$, in addition to a significant interaction between the two, Wilks’ $\lambda = .92, F(8, 560) = 3.04, p = .002, \eta_p^2 = .04$, on perceptions of the evidence quality. For all three, univariate tests indicated that the results were due to differences in perceptions of validity, rather than helpfulness. Article type significantly influenced ratings of evidence validity, $F(4, 281) = 10.65, p < .001, \eta_p^2 = .13$, but not of evidence helpfulness, $F(4, 283) = 0.89, p = .47, \eta_p^2 = .01$. Ratings of evidence validity were significantly higher for participants in the control group than for those who read the complex ($p < .001, M_{\text{diff}} = 0.91 [95\% \text{ CI: 0.41, 1.40}]$), simple ($p < .001, M_{\text{diff}} = 0.90 [95\% \text{ CI: 0.40, 1.39}]$), or story-and-fact articles ($p < .001, M_{\text{diff}} = 0.69 [95\% \text{ CI: 0.24, 1.14}]$), but did not significantly differ from those that read the story article ($p = .43, M_{\text{diff}} = 0.29 [95\% \text{ CI: -0.17, 0.74}]$) (for means see Table 20). Further, evidence validity ratings were significantly higher in the story group than they were in the complex ($p = .006, M_{\text{diff}} = 0.62 [95\%$
CI: 0.12, 1.12], and simple groups (p = .007, $M_{\text{diff}} = 0.61$ [95% CI: 0.11, 1.11]), although they did not significantly differ from the story-and-fact group (p = .10, $M_{\text{diff}} = 0.41$ [95% CI: -0.04, 0.86]). No other experimental groups differed significantly (all ps > .7).

Evidence type also significantly influenced ratings of evidence validity, $F(1, 281) = 31.37$, $p < .001$, $\eta_p^2 = .10$, but not helpfulness, $F(1, 281) = 1.74$, $p = .19$, $\eta_p^2 = .006$. Toolmark evidence ($M = 4.57$, $SD = 0.90$) was rated as significantly more valid than bitemark evidence ($M = 3.96$, $SD = 1.14$). The interaction between article and evidence type also was significant for ratings of evidence validity, $F(4, 281) = 4.26$, $p = .002$, $\eta_p^2 = .06$, but not helpfulness, $F(4, 281) = 0.68$, $p = .61$, $\eta_p^2 = .01$. For evidence validity, the articles influenced perceptions of both bitemark, $F(4, 138) = 10.98$, $p < .001$, $\eta_p^2 = .24$, and toolmark validity, $F(4, 143) = 2.50$, $p = .045$, $\eta_p^2 = .07$, but the effect was larger for bitemark evidence. Bitemark evidence validity ratings were higher for participants in the control group than those who read the complex ($p < .001, M_{\text{diff}} = 1.22$ [95% CI: 0.44, 2.00]), simple ($p < .001, M_{\text{diff}} = 1.25$ [95% CI: 0.49, 2.02]), or story-and-fact articles ($p < .001, M_{\text{diff}} = 1.32$ [95% CI: 0.61, 2.03]), but did not significantly differ from those that read the story article ($p = .69, M_{\text{diff}} = 0.33$ [95% CI: -0.37, 1.02]). Further, evidence validity ratings were significantly higher in the story group than they were in the complex ($p = .01, M_{\text{diff}} = 0.90$ [95% CI: 0.14, 1.65]), simple ($p = .006, M_{\text{diff}} = 0.93$ [95% CI: 0.19, 1.66]), and story-and-fact groups ($p = .001, M_{\text{diff}} = 0.99$ [95% CI: 0.31, 1.67]). No other differences between experimental groups reached significance (all ps > .9). Toolmark evidence was rated as marginally more valid in the control group than it was in the complex condition ($p = .06, M_{\text{diff}} = 0.63$ [95% CI: -0.01, 1.27]), but the control group did not differ from the simple ($p = .20, M_{\text{diff}} = 0.51$ [95% CI: -0.14, 1.17]), story ($p = .85, M_{\text{diff}} = 0.22$ [95% CI: -0.39, 0.83]), or
story-and-fact groups, \( (p = .97, M_{\text{diff}} = 0.14 [95\% \text{ CI: -0.44, 0.72}]) \). There were no differences between experimental groups (all \( ps > .2 \)).

**Perceived understanding of the expert and evidence.** Regarding perceived understanding, I found a significant main effect of article type on participants’ ratings of how well they understood the expert’s testimony, \( F(4, 281) = 2.90, \ p = .02, \eta^2_p = .04 \). There was no effect of evidence type, \( F(1, 281) = 2.60, \ p = .11, \eta^2_p = .009 \), and the interaction was not significant, \( F(4, 281) = 0.60, \ p = .66, \eta^2_p = .009 \). Follow-up analyses indicated that participants in the control group reported understanding the expert’s testimony better than participants who read the complex article \( (p = .01, M_{\text{diff}} = 0.94 [95\% \text{ CI: 0.16, 1.72}]) \) (for means see Table 19); no other differences reached significance (all \( ps > .1 \)). There was a significant effect of article type, \( F(4, 281) = 3.47, \ p = .009, \eta^2_p = .05 \), and a marginally significant effect of evidence type, \( F(1, 281) = 3.47, \ p = .06, \eta^2_p = .01 \), on participants’ ratings of how well they understood the forensic technique used by the expert; the interaction was not significant, \( F(4, 281) = 1.17, \ p = .33, \eta^2_p = .02 \). Follow-up analyses indicated that participants in the control group reported understanding the expert’s testimony better than participants who read the complex article \( (p = .02, M_{\text{diff}} = 0.99 [95\% \text{ CI: 0.09, 1.89}]) \) or the simple article \( (p = .01, M_{\text{diff}} = 1.06 [95\% \text{ CI: 0.16, 1.96}]) \); they did not differ from participants who read the story article \( (p = .20, M_{\text{diff}} = 0.65 [95\% \text{ CI: -0.18, 1.48}]) \) or the story-and-fact article \( (p = .33, M_{\text{diff}} = 0.56 [95\% \text{ CI: -0.25, 1.38}]) \) and no differences between experimental groups reached significance (all \( ps > .1 \)). Participants also reported understanding the toolmark evidence better \( (M = 6.87, SD = 1.65) \) than participants who viewed the bitemark evidence \( (M = 6.53, SD = 1.83) \).
Hypothesis 15: Results will differ for high and low NC participants. For low NC participants, beliefs about the validity of the evidence, perceptions of guilt, and perceptions of the expert and evidence will be highest in simple article conditions and lowest in story-and-fact article conditions, matching the overall pattern. For high NC participants, there will be no difference between the two fact-based article conditions, but both will be significantly higher than the story and story-and-fact article conditions.

My final hypothesis was not supported. There was no evidence that any of the articles had differential effects on low and high NC participants. Indeed, I found no influence of NC on validity beliefs, perceptions of guilt, or my primary measures of perceptions of the expert or evidence, and there were no significant interactions. NC influenced only perceived understanding of the expert’s testimony and his forensic technique. As in the Studies 1 and 2, only significant effects of NC or significant interactions with NC are mentioned in the analyses below. However, with the exception of validity beliefs, all models included NC group, article type, and evidence type.

**Pretrial validity beliefs.** An ANOVA revealed that there was no effect of NC on perceptions of the validity of bitemark evidence, $F(2, 275) = 0.50, p = .61, \eta^2_p = .004$, nor was there a significant interaction between NC and article type, $F(8, 275) = 1.02, p = .42, \eta^2_p = .03$. Similarly, there was no effect of NC on perceptions of the validity of other types of evidence, Wilks’ $\lambda = .93, F(16, 536) = 1.27, p = .21, \eta^2_p = .04$, and no interaction between NC and article type, Wilks’ $\lambda = .78, F(64, 1552) = 1.10, p = .28, \eta^2_p = .03$.

**Perceptions of guilt.** For verdicts, a logistic regression including NC, article type, evidence type, and all interactions with NC, was significant, $\chi^2(25) = 40.05, p = .03$, $R^2_{nag} = .18$, but NC was not a significant predictor of verdict choice, Wald $\chi^2(2) = 1.88, p = .39$, and no
interactions with NC reached significance (all $p$s $>.1$). The effect of article type became significant, Wald $\chi^2(4) = 10.14$, $p = .04$; however, given that the effect was not significant in the primary analysis and that follow-ups including only article type did not reveal any significant difference in verdicts, Wald $\chi^2(4) = 3.02$, $p = .56$, I did not consider this a reliable finding. There was no effect of NC on verdict confidence, $F(2, 260) = 0.39$, $p = .68$, $\eta_p^2 = .003$ and no interactions with NC reached significance (all $p$s $>.2$). There was also no effect of NC on estimates of the probability that the defendant murdered the victim, $F(2, 260) = 0.43$, $p = .65$, $\eta_p^2 = .003$, or that he was the source of the evidence, $F(2, 260) = 0.56$, $p = .57$, $\eta_p^2 = .004$, and no significant interactions with NC for either measure (all $p$s $>.1$).

**Perceptions of the expert and evidence.** There was no effect of NC on participants’ ratings of the credibility of the expert, $F(2, 260) = 0.77$, $p = .46$, $\eta_p^2 = .006$, and no interactions with NC reached significance (all $p$s $>.1$). There was also not a main effect of NC on participants’ ratings of evidence quality, Wilks’ $\lambda = .99$, $F(4, 518) = 0.42$, $p = .80$, $\eta_p^2 = .003$, and again no interactions with NC reached significance (all $p$s $>.065$).

For perceived understanding, there were significant effects of NC on participants’ ratings of how well they understood the expert’s testimony, $F(2, 260) = 11.29$, $p < .001$, $\eta_p^2 = .08$, and the forensic technique used by the expert, $F(2, 260) = 4.77$, $p = .009$, $\eta_p^2 = .04$, but no interactions with NC reached significance (all $p$s $>.2$). Additionally, the effect of evidence type on perceived understanding of the expert’s forensic technique was no longer even marginally significant, $F(1, 260) = 2.17$, $p = .14$, $\eta_p^2 = .008$. High NC participants said that they understood the expert significantly better than low ($p < .001, M_{\text{diff}} = 0.99$ [95% CI: 0.50, 1.49]) or moderate NC participants ($p = .001, M_{\text{diff}} = 0.74$ [95% CI: 0.25, 1.23]) (for means see Table 21); low and moderate NC participants did not differ ($p = .44, M_{\text{diff}} = 0.25$ [95% CI: -0.23, 0.73]). High NC
participants also said that they understood the forensic technique used by the expert significantly better than low NC participants ($p = .01$, $M_{\text{diff}} = 0.71$ [95% CI: 0.13, 1.30]) and marginally better than moderate NC participants ($p = .06$, $M_{\text{diff}} = 0.55$ [95% CI: -0.02, 1.13]); low and moderate NC participants again did not differ ($p = .79$, $M_{\text{diff}} = 0.16$ [95% CI: -0.41, 0.73]).
CHAPTER 17: STUDY THREE DISCUSSION

Contrary to my expectations, reading an article depicting a story about a wrongful conviction due to bitemark evidence was the least effective method of reducing both validity beliefs and reliance on forensic science evidence. These results are surprising, given that prior research has found a story to be an extremely persuasive tool (Green & Brock, 2005; Slater, 2002). Instead, the story had to be supplemented by factual information attacking the discipline in order for it to successfully influence participants’ perceptions of trial evidence. Indeed, the story-and-fact article was as effective as the complex article, and there was some evidence that it was more effective: only the story-and-fact article influenced estimates of the probability that the defendant was the source of the evidence. This finding is consistent with prior research showing that combining factual information with a narrative may increase its persuasive power (Slater & Rouner, 1996).

When only the story was presented, participants’ pretrial beliefs regarding the validity of bitemark evidence were only marginally lower than they were in the control group, and validity beliefs were higher than for those who had read the story-and-fact article or either fact-based article. Further, the story article did not influence perceptions of guilt or perceptions of the expert or evidence, even when other articles did so. The ineffectiveness of the story article in influencing perceptions of guilt is particularly surprising given that Greene and her colleagues (Greene & Loftus, 1984; Greene & Wade, 1988) found that reading about a wrongful conviction decreased guilty verdicts in a subsequent trial. Although the wrongful conviction described in their studies resulted from a misidentification, it is not clear why a wrongful conviction due to faulty forensic science would not have the same impact. It does suggest the possibility, however,
that forensic science—or beliefs in the probative value of forensic science—might be harder to challenge than other types of evidence.

In fact, the articles had almost no influence on perceptions of guilt. None of the articles significantly influenced verdicts, verdict confidence, or estimates of the probability that the defendant murdered the victim. However, as mentioned above, the story-and-fact article influenced estimates of the probability that the defendant was the source of bitemark evidence. That is, reading an article that combined a wrongful conviction story with facts attacking bitemark evidence significantly reduced bitemark source likelihood estimates compared to participants who read only unrelated articles. In contrast to Study 2, the complex article did not even marginally influence source estimates.

The story-based articles did not appear to be more likely to generalize. That is, I did not find that the story-based articles were more likely to influence perceptions of other forensic disciplines than the fact-based articles. In fact, there was some evidence that the complex article was more likely to generalize: when toolmark and bitemark evidence were examined separately, only the complex article even marginally reduced perceptions of the validity of the toolmark evidence compared to the control group. Further, the complex article significantly reduced pretrial validity-belief ratings of toolmark evidence (both in Study 2 and Study 3), while neither the story nor the story-and-fact article reduced ratings for any type of evidence other than bitemark. The complex article may have been more likely to generalize because it opened with information about the NAS Report’s criticisms of the forensic sciences more broadly. By contrast, the story-and-fact article only briefly mentioned that the NAS Report had found problems with the forensic sciences in general before targeting bitemark evidence. However, given that the complex and story-and-fact articles both influenced perceptions of the credibility
of the expert for both bitemark and toolmark evidence, further investigation of generalizability is needed.

I did not find that low and high NC participants were differentially influenced by the articles. Indeed, I found no differences between low and high NC participants on any of my perceptions of guilt measures or on ratings of the credibility of the expert or the quality of the evidence. Further, I did not replicate the Study 2 finding that high but not low NC participants differentiated between toolmark and bitemark evidence in their ratings of the credibility of the expert. However, there was some evidence from Study 2 that the affirming article most influenced low NC participants while the corrective article most influenced high NC participants. Neither of these articles was included in Study 3, and perhaps their absence could explain the failure to replicate (although I did not find differences between the articles for expert credibility in Study 2). I did replicate the finding from both Study 1 and Study 2 that high NC participants reported that they understood the expert’s testimony and the forensic technique upon which it was based better than low NC participants.

Taken as a whole, the results of Study 3 suggest that a story alone may be able to influence perceptions of the forensic sciences to a limited degree, but is not powerful enough to influence perceptions of such evidence if subsequently presented at trial. However, combining a story with factual information may be more persuasive than either presented on its own. That said, it is not clear to what extent combining a story with facts is more likely to generalize than facts presented without a narrative. Additional research on the specific factors that might increase generalizability is needed.
CHAPTER 18: GENERAL DISCUSSION

Consistent with previous research (Hans et al., 2011; Lieberman et al., 2008), participants viewed forensic science evidence as more valid and reliable than other types of evidence and their pretrial beliefs predicted post-trial perceptions of evidence quality. Further, results suggest that less valid forms of forensic science are given more weight than appropriate when they are presented as trial evidence: the forensic identification of a defendant using DNA, fingerprint, toolmark, or bitemark analysis all resulted in considerably more guilty verdicts than when no forensic evidence was presented. Participants did appear to have some sense of the comparative validity of different sciences, giving the highest evidence quality ratings to DNA evidence, but they appeared to overvalue the less valid types of evidence. Fingerprint evidence had a similar influence to DNA on perceptions of guilt, and no post-trial differences between toolmark and bitemark evidence were found, despite the fact that the NAS Report found fingerprint evidence to be less valid than DNA and bitemark evidence to be less valid than toolmark evidence.

Perhaps a greater problem is the fact that the forensic science evidence was viewed as better quality when the expert had testified that it matched the defendant than when he testified that it did not match. The expert was even rated as more credible when he testified that the evidence matched than when he said it did not. This result could not have been caused by differences in the testimony itself. Indeed, the testimony remained exactly the same whether the evidence matched or did not match with the exception of the expert’s ultimate conclusion. Further, in both cases the expert testified as to his conclusion to a reasonable degree of scientific certainty, ostensibly bolstering both a match and non-match equally (Garrett & Mitchell, 2013). It is not clear why the failure to find a match would reflect poorly on the expert or on his forensic technique. Indeed, the purpose of forensic analysis should be to test the possibility of a match
between evidence and a particular source, rather than confirming that they do match. However, it is possible that jurors in some way see it as a failure of the expert if he finds that they do not. An alternate possible cause would be that jurors view a non-match as having less probative value. However, participants did not report finding a non-match less helpful, so it does not seem that differences in perceived probative value alone can explain the differences in perceived quality. While replication of this finding is needed, it suggests that there may be some degree of prejudice to the side presenting an expert who says that evidence does not match a particular source, especially if the opposing side presents an expert who says that it does. If this finding was the result of the non-match being presented by a defense rather than a prosecution expert, it suggests direct prejudice to the defense.

Consistent with prior research (Kovera et al., 1999; Lieberman et al., 2008; McQuiston-Surrett & Saks, 2009), the traditional legal safeguard of cross-examination was not able to reduce the percentage of participants rendering guilty verdicts. Contrary to my predictions, however, having participants read articles that attacked the validity of bitemark evidence prior to reading trial materials also did not influence verdicts. The failure to influence verdicts did not mean that the articles had no impact: the articles were able to reduce pretrial perceptions of the validity of bitemark evidence and post-trial perceptions of the quality of bitemark and, sometimes, toolmark evidence. These results suggest that media campaigns designed to educate the general public about the deficiencies of forensic science may, in fact, help to correct some misconceptions about validity, but that they may not be sufficient to reduce the influence of expert testimony about a forensic match.

Regarding the most effective form of factual news reports, attacking the evidence from multiple angles using complex language was more effective than focusing on one particular
problem in greater detail and in simpler language. Supplementing a simple article with information that either changed the focus of the article to reinforce beliefs in the importance of the criminal justice system (affirming article) or that gave participants specific information about how to evaluate bitemark evidence at trial (corrective article) did not appear to have much additional influence on participants’ perceptions or judgments, and was still less effective than a complex article.

The results of Study 3 suggest that media campaigns that present only stories may not be an effective way of changing perceptions of the validity of forensic science evidence or trial judgments. It appears that the impact of such stories can be increased greatly, however, by supplementing the story with even one fact-based attack on the science reported in the story. Presenting a story along with a simple fact was just as effective as presenting a factual report that attacked forensic science from multiple angles, and sometime more effective—in Study 3, only the story-and-fact article influenced estimates of the likelihood that the defendant was the source of a bitemark found on the murder weapon. This finding suggests that the combination of narrative and factual information may be the most effective way in which to educate the public about science. The combination may also have greater persuasive power than either presented alone. If so, it suggests that attorneys may wish to use both facts and stories in making their arguments to the jury. This idea is also consistent with the story model (Pennington & Hastie, 1986, 1992), which indicates that arranging the evidence to fit a story may aid in comprehension and increase the effectiveness of an attorney’s argument. Further, this finding suggests that policy makers wishing to communicate information to the public or to drum up support for a particular measure might be well served by presenting facts alongside an illustrative story.
Need for cognition does not appear to consistently influence jurors’ perceptions and judgments of forensic science evidence with the exception of perceived understanding. Across all three studies, high NC participants reported the greatest understanding of the expert and his forensic technique. It is possible that their greater perceived understanding may help to explain the Study 1 finding that high NC participants appeared to give greater weight to a forensic match than low NC participants (see also Garrett & Mitchell’s [2013] similar finding with numeracy). If high NC participants felt that they had a better grasp on what the expert said and the manner in which he conducted his analyses, they may have been more likely to accept automatically that his conclusions regarding source were correct (Gilbert et al., 1990, 1993). That said, I did not measure actual comprehension of the expert’s testimony, so it is not clear whether high and low NC participants differed in their actual understanding of the evidence or whether high NC participants only felt that they understood it better.

**Limitations**

There were several limitations to these studies. First, all three studies relied on online samples and presented stimulus materials online. While the use of participants recruited on MTurk has been validated in prior research (Buhrmester et al., 2011; Crump et al., 2013; Paolacci et al., 2010; Sprouse, 2011), having people participate online could still be problematic since it is difficult to monitor the extent of individual participants’ attention to stimulus materials. Further, it is not possible to control for reading speed, or for the possibility that some participants may have read individual pages of the transcript multiple times prior to moving to the next page. To minimize the former limitation, I only retained participants if they passed a number of attention/manipulation checks. While it was not possible to control for the latter, the same problems would have been encountered with an in-person sample since participants would
be given transcripts to read on their own. Of course, these problems could have been avoided by using a videotaped trial rather than written trial transcripts; thus replication of my results using a videotaped trial is needed. However, with one exception (Lieberman et al., 2008, Study 3), the limited body of research on jurors’ perceptions of forensic evidence has used trial summaries (e.g., Garrett & Mitchell, 2013; Lieberman et al., 2008, Studies 1 and 2; McQuiston-Surrett & Saks, 2008, 2009), or has presented only the expert’s testimony in transcript form (Koehler, 2011); thus, the use of a full trial transcript represents an increase in ecological validity compared to most prior research in the area.

A limitation of Studies 2 and 3 was that no delay separated the first part of the studies, reading the articles, from the second part, reading the trial transcript. Further, it is possible that some participants did not believe the cover story, and figured out that the two parts were actually part of one larger study. These limitations are a particular problem given that these studies investigated whether media campaigns designed to educate the general public could also influence jurors in subsequent trials. However, the use of MTurk participants made it extremely difficult to impose a uniform period of delay for all participants; further, it is likely that I would have lost many more participants to attrition. To attempt to minimize these limitations, I had experimental participants read two unrelated articles on criminal justice topics. While it is possible that a larger number of articles would have provided better cover, it is likely that a greater number of participants would not have finished the study due to the increased time involved. Further, it is likely that more participants would have been lost due to failure to attend to the materials if the amount that they were required to read was increased greatly. Finally, given that no previous research examining whether media reports can influence juror evaluations of forensic evidence have been conducted, increasing the signal-to-noise ratio was preferable in
order to see whether any effects could be found. Future research should investigate whether the same results would be found when a delay was included and when the target article was embedded in a larger number of unrelated articles.

**Future Directions**

Despite its limitations, the present research has demonstrated the persuasiveness of forensic science evidence and the difficulty of countering such evidence. It also suggests several different avenues for future research. Such research could focus both on the trial process itself and on other forms of media campaigns designed to educate the public about unreliable evidence.

**Replication.** First, it is necessary to replicate the results of these studies in order to determine that the results are robust and can be generalized to other populations and types of cases. Further, the sheer number of analyses conducted increases the potential for Type II error; thus, replication of these results would increase the confidence with which conclusions can be drawn from them. Additionally, such research could address some of the limitations outlined above by using off-line samples drawn from community samples or real jury pools, and using longer transcripts or a full trial video. In particular, direct replication of this study in other laboratories using other samples and other stimuli would enhance reliability and allow for tests of potential moderators or situational influences on these results (Pashler & Harris, 2012; Simons, 2014).

**Identification versus non-identification.** It was not possible to determine from these studies whether non-matching evidence—and the credibility of the expert presenting it—were lower than matching evidence simply because the evidence did not match or because the non-match was presented by a defense, rather than prosecution, expert. The former suggests that
either side could incur prejudice if their expert testifies about a non-match, while the latter suggests that the prejudice is specific to the defense. Thus, future research manipulating match type and the side presenting the match versus non-match is needed.

Legal safeguards. Another avenue for future research is to investigate the effectiveness of other legal safeguards in addition to cross-examination, such as judicial instructions or opposing expert witnesses. Examination of whether using multiple safeguards might have an additive effect could be particularly fruitful. Cross-examination was not effective in influencing verdicts; similarly, prior research suggests that judicial instructions (e.g., McQuiston-Surrett & Saks, 2009) or opposing experts (e.g., Levett & Kovera, 2008) may not be sufficient to counteract unreliable evidence. Nevertheless, cross-examination was sufficient to impact perceptions of the evidence and the expert himself, and prior research has also found that it may influence perceived evidence strength (Koehler, 2011) or understanding (McQuiston-Surrett & Saks, 2009). It is possible that if the cross-examination were reinforced by testimony from an opposing expert or judicial instructions that specifically targeted the forensic science, the combination would be powerful enough to counter the direct testimony of a forensic expert to such an extent that verdicts might be influenced. If the judicial instructions were presented prior to the expert’s testimony, the combination might be particularly powerful.

Media campaigns. The present research represented only an initial step in investigating the types of media campaigns that might educate the general public, and also make them better evaluators of forensic evidence if called as jurors. None of the articles was able to influence verdicts, thus research on how to strengthen media campaigns is needed. It is not clear what specifically made the complex article the most effective fact-based article. While it may have been the presence of multiple criticisms of bitemark evidence, the complex article also used
more advanced and sophisticated language and either could have influenced participants. Indeed, prior research suggests that people might be more convinced by language that sounds more “scientific,” regardless of the content of such language (e.g., Weisberg, Keil, Goodstein, Rawson, & Gray, 2008). Thus, it is possible that variations of the complex report, rather than the simple report, that either affirmed people’s belief in the criminal justice system or that supplemented the criticisms with information about how unreliable science should be evaluated might be particularly effective. Given the evidence that the report that combined facts with an illustrative story may have had the greatest influence, however, it is possible that variations of the story-and-fact article that combined elements of the complex article with the story might be the best way to maximize the effectiveness of educational media. For example, it is possible that supplementing the story with more or other types of facts might have a greater impact. Modifying the framing of the story or adding information about how to evaluate trial evidence—particularly if it explicitly connected the story to future trials—might also be effective.

Further research on what types of wrongful conviction stories might be most influential is also needed. Previous research found that reading about a wrongful conviction due to misidentification decreased guilty verdicts in subsequent trials (Greene & Loftus, 1984; Greene & Wade, 1988), while I found that reading about a wrongful conviction due to faulty forensic evidence did not. These results suggest the possibility that information about different contributing causes to wrongful convictions may have differential impacts. Beliefs about forensic science might be especially difficult to challenge particularly when compared to eyewitness evidence: it may be considerably easier to accept that people are often wrong than that science is often wrong. Indeed, both the present research and previous research (Hans et al., 2011; Lieberman et al., 2008) suggest that people generally view eyewitness evidence with greater
skepticism than other types of evidence, particularly forensic evidence. It is also possible, then, that media campaigns targeting unreliable evidence are more effective when the evidence is already viewed with greater skepticism. Such an explanation would make sense intuitively; however, the prior PTP research was conducted in the 1980s, and it is not clear to what extent the general public was aware of potential problems with eyewitness evidence at the time. Stories depicting wrongful convictions due to misidentification in the news media, documentaries, and even big-budget movies, have proliferated in recent years. Thus, new research investigating the influence of wrongful conviction media examining different contributing causes is needed.

**Conclusions**

Although additional research is needed, the present studies provide important information about the power of forensic science evidence. The results of this research suggest that although people may differentiate to some extent between less and more valid types of evidence, they may still be overly persuaded by forensic science evidence relative to its actual probative value. Although jurors can be induced to view the evidence as less valid using either cross-examination or pretrial media campaigns, neither appear to reduce their reliance on forensic science evidence to the extent that they are less likely to find a defendant guilty once a forensic identification has been made. Further, these results suggest that the extent to which jurors believe that a particular type of forensic science is valid when they enter the trial, regardless of whether their beliefs are naïve or informed by the media, may not determine their reliance on such evidence. Therefore, it is possible that the burden should be on judges to reject invalid forensic science evidence rather than on jurors to attempt to evaluate such complex evidence appropriately. However, the fact that participants’ perceptions of the evidence and expert were influenced by the experimental manipulations in the current research leaves some hope that strengthening attacks on forensic
science, both pretrial and during the trial, may eventually be able to lessen its powerful influence on jurors.
Table 1

Comparisons between jury and student samples on mean ratings for the accuracy and persuasiveness of different evidence types.

<table>
<thead>
<tr>
<th>Evidence Type</th>
<th>Student (n = 383)</th>
<th>Jury (n = 233)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA evidence of human fluid/tissue</td>
<td>93.7% (11.2)</td>
<td>94.9% (8.7)</td>
</tr>
<tr>
<td>Fingerprint evidence</td>
<td>90.1% (14.7)</td>
<td>91.4% (13.0)</td>
</tr>
<tr>
<td>Hair and fiber evidence</td>
<td>88.7% (16.1)</td>
<td>89.2% (13.8)</td>
</tr>
<tr>
<td>Video surveillance pictures</td>
<td>88.1% (16.1)</td>
<td>87.3% (15.9)</td>
</tr>
<tr>
<td>Alcohol and drug tests</td>
<td>79.2% (21.4)</td>
<td>78.4% (22.2)</td>
</tr>
<tr>
<td>Expert testimony by scientists</td>
<td>81.5% (18.1)</td>
<td>76.7% (19.5)**</td>
</tr>
<tr>
<td>Suspect’s confession</td>
<td>73.7% (25.3)</td>
<td>76.4% (22.9)</td>
</tr>
<tr>
<td>Victim statements in testimony</td>
<td>66.8% (20.5)</td>
<td>70.0% (19.4)</td>
</tr>
<tr>
<td>Eyewitness statements in testimony</td>
<td>62.8% (21.1)</td>
<td>69.7% (19.3)**</td>
</tr>
<tr>
<td>Persuasiveness of evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspect’s DNA found on victim</td>
<td>92.6% (13.0)</td>
<td>94.2% (9.1)</td>
</tr>
<tr>
<td>Suspect’s DNA found at crime scene</td>
<td>89.5% (15.6)</td>
<td>89.9% (13.5)</td>
</tr>
<tr>
<td>Suspect confessed to the crime</td>
<td>88.9% (16.0)</td>
<td>86.4% (18.1)</td>
</tr>
<tr>
<td>Suspect’s hair fibers found at scene</td>
<td>82.2% (20.3)</td>
<td>83.5% (17.5)</td>
</tr>
<tr>
<td>Positive ID of suspect by victim</td>
<td>75.5% (21.2)</td>
<td>78.9% (17.9)*</td>
</tr>
<tr>
<td>Positive ID of suspect by eyewitness</td>
<td>69.8% (22.1)</td>
<td>73.3% (20.0)*</td>
</tr>
</tbody>
</table>

Note. Standard deviation presented in parentheses. ID = identification
* p < .05. ** p < .01.

Table 2

Study 1. Percentage (n) of participants agreeing with individual items comprising beliefs-in-science and reservations-about-science composite variables.

<table>
<thead>
<tr>
<th>Beliefs-in-science</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science and technology are making our lives healthier, easier, and more comfortable.</td>
<td>0.4% (3)</td>
<td>3.9% (27)</td>
<td>52.9% (367)</td>
<td>42.8% (297)</td>
</tr>
<tr>
<td>Most scientists want to work on things that will make life better for the average person.</td>
<td>0.1% (1)</td>
<td>6.3% (44)</td>
<td>62.8% (436)</td>
<td>30.7% (213)</td>
</tr>
<tr>
<td>With the application of science and technology, work will become more interesting.</td>
<td>0.9% (6)</td>
<td>12.8% (89)</td>
<td>57.3% (398)</td>
<td>29.0% (201)</td>
</tr>
<tr>
<td>Because of science and technology, there will be more opportunities for the next generation.</td>
<td>0.6% (4)</td>
<td>11.7% (81)</td>
<td>47.4% (329)</td>
<td>40.2% (279)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reservations-about-science</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>We depend too much on science and not enough on faith.</td>
<td>40.6% (282)</td>
<td>38.8% (269)</td>
<td>15.3% (105)</td>
<td>5.2% (36)</td>
</tr>
<tr>
<td>It is not important for me to know about science in my daily life.</td>
<td>38.6% (268)</td>
<td>42.5% (295)</td>
<td>14.1% (98)</td>
<td>4.6% (32)</td>
</tr>
<tr>
<td>Science makes our way of life change too fast.</td>
<td>27.2% (189)</td>
<td>51.4% (357)</td>
<td>17.9% (124)</td>
<td>3.3% (23)</td>
</tr>
</tbody>
</table>
Table 3

Study 1. Reliability ratings for each type of evidence on a 1-5 point scale and results of paired-sample t-tests for each evidence comparison and number (percentage) of participants saying they have never heard of the type of evidence in lieu of rating.

<table>
<thead>
<tr>
<th></th>
<th>Fingerprint</th>
<th>Toolmark</th>
<th>Bitemark</th>
<th>Eyewitness</th>
<th>Expert</th>
<th>Police</th>
<th>Victim</th>
<th>Confession</th>
<th>Polygraph</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DNA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 4.75</td>
<td>t (670) =</td>
<td>t (486) =</td>
<td>t (621) =</td>
<td>t (672) =</td>
<td>t (673) =</td>
<td>t (655) =</td>
<td>t (673) =</td>
<td>t (672) =</td>
<td>16 (2.3%)</td>
<td></td>
</tr>
<tr>
<td>SD = 0.50</td>
<td>15.24</td>
<td>29.84</td>
<td>22.23</td>
<td>35.77</td>
<td>40.55</td>
<td>37.38</td>
<td>38.71</td>
<td>52.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fingerprint</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 4.33</td>
<td>t (490) =</td>
<td>t (623) =</td>
<td>t (676) =</td>
<td>t (673) =</td>
<td>t (674) =</td>
<td>t (654) =</td>
<td>t (673) =</td>
<td>t (673) =</td>
<td>16 (2.3%)</td>
<td></td>
</tr>
<tr>
<td>SD = 0.77</td>
<td>17.03</td>
<td>11.67</td>
<td>38.04</td>
<td>21.79</td>
<td>25.25</td>
<td>23.90</td>
<td>25.24</td>
<td>41.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Toolmark</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 3.58</td>
<td>t (497) =</td>
<td>t (500) =</td>
<td>t (499) =</td>
<td>t (488) =</td>
<td>t (498) =</td>
<td>t (501) =</td>
<td>t (633) =</td>
<td>190 (27.4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 0.86</td>
<td>-7.75</td>
<td>2.24</td>
<td>3.12</td>
<td>5.26</td>
<td>18.95</td>
<td>18.95</td>
<td>27.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p &lt; .001</td>
<td>p = .03</td>
<td>p = .02</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bitemark</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 3.89</td>
<td>t (634) =</td>
<td>t (629) =</td>
<td>t (632) =</td>
<td>t (612) =</td>
<td>t (631) =</td>
<td>t (632) =</td>
<td>t (633) =</td>
<td>58 (8.4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 0.96</td>
<td>22.86</td>
<td>10.30</td>
<td>10.57</td>
<td>12.43</td>
<td>27.95</td>
<td>27.95</td>
<td>27.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eyewitness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 2.84</td>
<td>t (684) =</td>
<td>t (687) =</td>
<td>t (665) =</td>
<td>t (686) =</td>
<td>t (686) =</td>
<td>t (685) =</td>
<td>7 (1%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 0.93</td>
<td>-16.71</td>
<td>-16.84</td>
<td>-12.08</td>
<td>6.57</td>
<td>21.01</td>
<td>21.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Expert</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 3.51</td>
<td>t (683) =</td>
<td>t (663) =</td>
<td>t (682) =</td>
<td>t (682) =</td>
<td>t (682) =</td>
<td>4 (0.6%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 0.83</td>
<td>1.08</td>
<td>1.42</td>
<td>4.34</td>
<td>1.08</td>
<td>21.01</td>
<td>21.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p = .28</td>
<td>p = .16</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Police</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 3.47</td>
<td>t (664) =</td>
<td>t (684) =</td>
<td>t (684) =</td>
<td>t (662) =</td>
<td>t (684) =</td>
<td>26 (3.7%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 0.77</td>
<td>0.69</td>
<td>3.78</td>
<td>3.78</td>
<td>2.71</td>
<td>19.33</td>
<td>19.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p = .49</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p = .007</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Victim</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 3.45</td>
<td>t (664) =</td>
<td>t (664) =</td>
<td>t (664) =</td>
<td>t (662) =</td>
<td>t (684) =</td>
<td>4 (0.6%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 0.86</td>
<td>0.69</td>
<td>2.71</td>
<td>2.71</td>
<td>16.85</td>
<td>16.85</td>
<td>16.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p = .49</td>
<td>p = .007</td>
<td>p = .007</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Confession</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 3.32</td>
<td>t (684) =</td>
<td>t (684) =</td>
<td>t (684) =</td>
<td>t (684) =</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 0.94</td>
<td>0.69</td>
<td>3.78</td>
<td>3.78</td>
<td>3.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p = .49</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Polygraph</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 2.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 1.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4

Study 1. Percentage (n) of guilty verdicts in different evidence type conditions dependent on whether the expert testified that the defendant was the source of the evidence (match) or not the source of the evidence (non-match).*

<table>
<thead>
<tr>
<th>Evidence Type</th>
<th>Non-Match (N = 330)</th>
<th>Match (N = 313)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>6.1% (5)a</td>
<td>79.0% (64)b</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>7.7% (6)a</td>
<td>68.8% (55)b</td>
</tr>
<tr>
<td>Toolmark</td>
<td>11.1% (9)a</td>
<td>51.4% (35)c</td>
</tr>
<tr>
<td>Bitemark</td>
<td>12.4% (11)a</td>
<td>51.3% (40)c</td>
</tr>
</tbody>
</table>

*Percentages with different superscripts significantly different (p < .05).
Table 5

*Study 1. Mean (SD) verdict confidence (1-7 scale) in different evidence conditions and when the evidence matched versus did not match the defendant.*

<table>
<thead>
<tr>
<th>Evidence Type</th>
<th>Confidence in Verdict</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>5.18&lt;sup&gt;a&lt;/sup&gt; (1.76)</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>5.01&lt;sup&gt;ab&lt;/sup&gt; (1.63)</td>
</tr>
<tr>
<td>Toolmark</td>
<td>4.54&lt;sup&gt;b&lt;/sup&gt; (1.79)</td>
</tr>
<tr>
<td>Bitemark</td>
<td>4.62&lt;sup&gt;b&lt;/sup&gt; (1.64)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Match Type</th>
<th>Confidence in Verdict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match</td>
<td>5.07&lt;sup&gt;a&lt;/sup&gt; (1.69)</td>
</tr>
<tr>
<td>Non-match</td>
<td>4.62&lt;sup&gt;b&lt;/sup&gt; (1.73)</td>
</tr>
</tbody>
</table>

*For each independent variable, means with different superscripts significantly different (p < .05).
<table>
<thead>
<tr>
<th>Evidence Type</th>
<th>Murder Probability Estimates</th>
<th>Source Probability Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Match</td>
<td>Match</td>
</tr>
<tr>
<td>DNA</td>
<td>41.00(^a) (26.62)</td>
<td>84.07(^a) (20.55)</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>42.96(^a) (26.45)</td>
<td>76.84(^{ab}) (23.98)</td>
</tr>
<tr>
<td>Toolmark</td>
<td>47.44(^a) (23.80)</td>
<td>73.08(^b) (22.07)</td>
</tr>
<tr>
<td>Bitemark</td>
<td>44.41(^a) (22.93)</td>
<td>70.99(^b) (22.81)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cross-examination Type</th>
<th>Murder Probability Estimates</th>
<th>Source Probability Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Match</td>
<td>Match</td>
</tr>
<tr>
<td>Evidence-focused</td>
<td>46.27(^a) (24.99)</td>
<td>75.54(^a) (25.72)</td>
</tr>
<tr>
<td>Expert-focused</td>
<td>43.73(^a) (24.34)</td>
<td>77.58(^a) (20.44)</td>
</tr>
<tr>
<td>Non-substantive</td>
<td>41.77(^a) (25.53)</td>
<td>75.92(^a) (22.35)</td>
</tr>
</tbody>
</table>

Overall 43.96 (24.95) 76.36 (22.84) 19.05 (22.31) 87.03 (17.41)

*For evidence type and cross-examination type, means within match and non-match conditions with different superscripts significantly different (p < .05). Match and non-match differ significantly for both types of probability estimate.
Table 7

Study 1. Mean (SD) ratings of the credibility of the expert (1-7) and of the validity and helpfulness of the evidence (1-6) in addition to participants’ perceptions of how well they understood the evidence (1-7) dependent on the type of evidence, whether the evidence matched the defendant, and the type of cross-examination used.

<table>
<thead>
<tr>
<th>Evidence Type</th>
<th>Expert Credibility</th>
<th>Evidence Validity</th>
<th>Evidence Helpfulness</th>
<th>Evidence Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>6.23ᵃ (0.96)</td>
<td>5.25ᵃ (0.77)</td>
<td>4.94ᵃ (1.29)</td>
<td>5.59ᵃᵇ (1.43)</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>6.09ᵃᵇ (0.88)</td>
<td>4.98ᵇ (0.88)</td>
<td>4.76ᵃᵇ (1.28)</td>
<td>5.89ᵃ (1.17)</td>
</tr>
<tr>
<td>Toolmark</td>
<td>5.81ᶜ (0.91)</td>
<td>4.69ᶜ (0.82)</td>
<td>4.46ᵇᶜ (1.11)</td>
<td>5.50ᵇ (1.27)</td>
</tr>
<tr>
<td>Bitemark</td>
<td>5.82ᵇᶜ (0.99)</td>
<td>4.64ᶜ (0.86)</td>
<td>4.30ᶜ (1.31)</td>
<td>5.60ᵃᵇ (1.27)</td>
</tr>
<tr>
<td>Match Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match</td>
<td>6.10ᵃ (0.95)</td>
<td>4.99ᵃ (0.87)</td>
<td>4.72ᵃ (1.28)</td>
<td>5.83ᵃ (1.17)</td>
</tr>
<tr>
<td>Non-match</td>
<td>5.88ᵇ (0.94)</td>
<td>4.80ᵇ (0.85)</td>
<td>4.52ᵃ (1.26)</td>
<td>5.47ᵇ (1.33)</td>
</tr>
<tr>
<td>Cross-examination Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidence-focused</td>
<td>5.74ᵃ (1.02)</td>
<td>4.59ᵃ (0.99)</td>
<td>4.57ᵃ (1.24)</td>
<td>5.48ᵃ (1.33)</td>
</tr>
<tr>
<td>Expert-focused</td>
<td>5.98ᵇ (1.02)</td>
<td>5.00ᵇ (0.76)</td>
<td>4.68ᵃ (1.30)</td>
<td>5.65ᵃᵇ (1.27)</td>
</tr>
<tr>
<td>Non-substantive</td>
<td>6.25ᵃ (0.73)</td>
<td>5.08ᵇ (0.74)</td>
<td>4.60ᵃ (1.28)</td>
<td>5.80ᵇ (1.26)</td>
</tr>
</tbody>
</table>

*For each independent variable, means within each DV with different superscripts significantly different (p < .05).
Study 1. Mean (SD) pretrial beliefs-in-science (4-16) and reservations-about-science ratings (3-12) for low, moderate, and high NC participants.

<table>
<thead>
<tr>
<th>NC Group</th>
<th>Beliefs-in-science</th>
<th>Reservations-about-science</th>
</tr>
</thead>
<tbody>
<tr>
<td>High NC</td>
<td>13.54&lt;sup&gt;a&lt;/sup&gt; (2.03)</td>
<td>4.95&lt;sup&gt;a&lt;/sup&gt; (1.75)</td>
</tr>
<tr>
<td>Moderate NC</td>
<td>12.89&lt;sup&gt;b&lt;/sup&gt; (1.80)</td>
<td>5.79&lt;sup&gt;b&lt;/sup&gt; (1.78)</td>
</tr>
<tr>
<td>Low NC</td>
<td>12.79&lt;sup&gt;b&lt;/sup&gt; (1.78)</td>
<td>6.19&lt;sup&gt;c&lt;/sup&gt; (1.72)</td>
</tr>
</tbody>
</table>

<sup>Within each DV, means with different superscripts significantly different (p < .05).</sup>
Table 9

*Study 1. Mean (SD) confidence in verdict choice (1-7), and ratings of the credibility of the expert (1-7), the validity and helpfulness of the evidence (1-6), and perceived understanding of the expert’s testimony and his forensic technique (1-7) for low, moderate and high NC participants.*

<table>
<thead>
<tr>
<th>NC Group</th>
<th>Verdict Confidence</th>
<th>Expert Credibility</th>
<th>Evidence Validity</th>
<th>Evidence Helpfulness</th>
<th>Expert Understanding</th>
<th>Evidence Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>High NC</td>
<td>5.01² (1.80)</td>
<td>6.25² (0.87)</td>
<td>5.12² (0.86)</td>
<td>4.78² (1.34)</td>
<td>6.51² (0.79)</td>
<td>5.96² (1.11)</td>
</tr>
<tr>
<td>Moderate NC</td>
<td>4.70¹ (1.71)</td>
<td>5.91ᵇ (0.94)</td>
<td>4.84ᵇ (0.81)</td>
<td>4.64ᵃᵇ (1.24)</td>
<td>6.11ᵇ (1.05)</td>
<td>5.64ᵇ (1.29)</td>
</tr>
<tr>
<td>Low NC</td>
<td>4.67ᵃ (1.71)</td>
<td>5.86ᵇ (0.99)</td>
<td>4.77ᵇ (0.86)</td>
<td>4.47ᵇ (1.22)</td>
<td>5.95ᵇ (1.11)</td>
<td>5.31ᶜ (1.32)</td>
</tr>
</tbody>
</table>

*Within each DV, means with different superscripts significantly different (p < .05).*
Table 10

*Study 1. Mean (SD) estimates of the probability that the defendant was the source of the evidence for low, moderate, and high NC participants dependent on whether the evidence matched or did not match the defendant.*

<table>
<thead>
<tr>
<th>NC group</th>
<th>Non-Match</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>High NC</td>
<td>13.41a (19.84)</td>
<td>90.56a (16.57)</td>
</tr>
<tr>
<td>Moderate NC</td>
<td>21.58b (24.08)</td>
<td>86.38a (18.36)</td>
</tr>
<tr>
<td>Low NC</td>
<td>21.64b (22.93)</td>
<td>86.99a (16.03)</td>
</tr>
</tbody>
</table>

*Means within match and non-match conditions with different superscripts significantly different (p < .05).*
Table 11

Study 2. Validity-belief ratings for control participants only for each type of evidence and results of paired-sample t-tests for each evidence comparison and average knowledge ratings for each evidence type.

<table>
<thead>
<tr>
<th>Validity-belief</th>
<th>Fingerprint</th>
<th>Toolmark</th>
<th>Bitemark</th>
<th>Eyewitness</th>
<th>Confession</th>
<th>Polygraph</th>
<th>Toxicology</th>
<th>Arson</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 89.83</td>
<td></td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td></td>
<td>M = 59.76</td>
</tr>
<tr>
<td>SD = 17.39</td>
<td></td>
<td>4.75</td>
<td>12.72</td>
<td>9.49</td>
<td>18.47</td>
<td>13.98</td>
<td>12.43</td>
<td></td>
<td>SD = 29.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fingerprint</td>
<td></td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td></td>
<td>M = 59.03</td>
</tr>
<tr>
<td>M = 80.60</td>
<td></td>
<td>9.15</td>
<td>7.23</td>
<td>14.58</td>
<td>10.58</td>
<td>9.98</td>
<td>-0.29</td>
<td></td>
<td>SD = 29.26</td>
</tr>
<tr>
<td>SD = 21.94</td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &gt; .001</td>
<td>p &gt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toolmark</td>
<td></td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td></td>
<td>M = 26.28</td>
</tr>
<tr>
<td>M = 55.97</td>
<td></td>
<td>5.81</td>
<td>2.46</td>
<td>2.07</td>
<td>2.07</td>
<td>9.42</td>
<td>-3.35</td>
<td></td>
<td>SD = 26.72</td>
</tr>
<tr>
<td>SD = 22.86</td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitemark</td>
<td></td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td></td>
<td>M = 34.68</td>
</tr>
<tr>
<td>M = 62.23</td>
<td></td>
<td>5.16</td>
<td>2.07</td>
<td>2.07</td>
<td>2.07</td>
<td>9.42</td>
<td>-3.35</td>
<td></td>
<td>SD = 29.28</td>
</tr>
<tr>
<td>SD = 23.96</td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eyewitness</td>
<td></td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td></td>
<td>M = 49.12</td>
</tr>
<tr>
<td>M = 39.18</td>
<td></td>
<td>9.48</td>
<td>3.70</td>
<td>6.18</td>
<td>6.18</td>
<td>9.41</td>
<td>-9.41</td>
<td></td>
<td>SD = 28.83</td>
</tr>
<tr>
<td>SD = 20.18</td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confession</td>
<td></td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td></td>
<td>M = 54.46</td>
</tr>
<tr>
<td>M = 47.81</td>
<td></td>
<td>3.30</td>
<td>3.30</td>
<td>3.30</td>
<td>3.30</td>
<td>3.30</td>
<td>3.30</td>
<td></td>
<td>SD = 29.27</td>
</tr>
<tr>
<td>SD = 23.73</td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polygraph</td>
<td></td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td></td>
<td>M = 48.88</td>
</tr>
<tr>
<td>M = 48.40</td>
<td></td>
<td>-1.9</td>
<td>-1.9</td>
<td>-1.9</td>
<td>-1.9</td>
<td>-1.9</td>
<td>-1.9</td>
<td></td>
<td>SD = 29.28</td>
</tr>
<tr>
<td>SD = 27.02</td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toxicology</td>
<td></td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td>t (67) =</td>
<td></td>
<td>M = 44.88</td>
</tr>
<tr>
<td>M = 81.21</td>
<td></td>
<td>8.36</td>
<td>8.36</td>
<td>8.36</td>
<td>8.36</td>
<td>8.36</td>
<td>8.36</td>
<td></td>
<td>SD = 29.32</td>
</tr>
<tr>
<td>SD = 19.58</td>
<td></td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arson</td>
<td></td>
<td>M = 64.07</td>
<td>M = 64.07</td>
<td>M = 64.07</td>
<td>M = 64.07</td>
<td>M = 64.07</td>
<td>M = 64.07</td>
<td></td>
<td>M = 33.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD = 20.09</td>
<td>SD = 20.09</td>
<td>SD = 20.09</td>
<td>SD = 20.09</td>
<td>SD = 20.09</td>
<td>SD = 20.09</td>
<td></td>
<td>SD = 28.72</td>
</tr>
</tbody>
</table>
Table 12

Study 2. *Mean (SD) validity-belief ratings in the different article conditions for bitemark and other evidence types.*

<table>
<thead>
<tr>
<th>Evidence Type</th>
<th>Control</th>
<th>Complex</th>
<th>Simple</th>
<th>Affirming</th>
<th>Corrective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitemark</td>
<td>62.23 (23.96)</td>
<td>24.48* (24.67)</td>
<td>32.68* (24.35)</td>
<td>34.00* (26.11)</td>
<td>33.48* (23.82)</td>
</tr>
<tr>
<td>DNA</td>
<td>89.93 (17.39)</td>
<td>85.83 (21.19)</td>
<td>91.48 (12.04)</td>
<td>86.62 (19.79)</td>
<td>90.10 (19.55)</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>80.60 (21.94)</td>
<td>74.28 (24.42)</td>
<td>75.22 (24.29)</td>
<td>70.86 (24.98)</td>
<td>75.22 (22.45)</td>
</tr>
<tr>
<td>Toolmark</td>
<td>55.97 (22.86)</td>
<td>37.93*† (25.18)</td>
<td>39.56* (24.96)</td>
<td>50.57† (23.07)</td>
<td>48.58 (23.23)</td>
</tr>
<tr>
<td>Eyewitness</td>
<td>39.18 (20.18)</td>
<td>31.09 (21.44)</td>
<td>33.58 (23.84)</td>
<td>40.35 (22.74)</td>
<td>37.15 (23.30)</td>
</tr>
<tr>
<td>Confession</td>
<td>47.81 (23.73)</td>
<td>41.09 (27.46)</td>
<td>46.19 (29.40)</td>
<td>50.36 (24.51)</td>
<td>46.04 (26.32)</td>
</tr>
<tr>
<td>Polygraph</td>
<td>48.40 (27.02)</td>
<td>37.18 (26.51)</td>
<td>30.74* (22.84)</td>
<td>34.77* (23.83)</td>
<td>37.62 (26.11)</td>
</tr>
<tr>
<td>Toxicology</td>
<td>81.21 (19.58)</td>
<td>70.22 (23.97)</td>
<td>71.27† (24.95)</td>
<td>75.49 (20.25)</td>
<td>70.94† (23.74)</td>
</tr>
<tr>
<td>Arson</td>
<td>64.07 (20.09)</td>
<td>51.74* (24.26)</td>
<td>51.28* (25.77)</td>
<td>56.14 (19.95)</td>
<td>55.40 (21.18)</td>
</tr>
</tbody>
</table>

*Significantly different from control group at $p < .05$.
†Different from control group at $p < .08$.
‡Within evidence types, article means significantly differ from each other.
Study 2. Mean (SD) ratings of the probability that the defendant was the source of the evidence (1-100), the credibility of the expert (1-9), and the validity and helpfulness of the evidence (1-6) dependent on which article was read.

<table>
<thead>
<tr>
<th>Article type</th>
<th>Source Probability</th>
<th>Expert Credibility</th>
<th>Evidence Validity</th>
<th>Evidence Helpfulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>80.92 (18.02)</td>
<td>7.64 (1.22)</td>
<td>4.91 (0.71)</td>
<td>4.18 (1.34)</td>
</tr>
<tr>
<td>Complex</td>
<td>70.25† (21.05)</td>
<td>6.99* (1.43)</td>
<td>3.69* (1.11)</td>
<td>3.60† (1.31)</td>
</tr>
<tr>
<td>Simple</td>
<td>73.06 (24.37)</td>
<td>7.38 (1.27)</td>
<td>3.93* (1.01)</td>
<td>3.82 (1.18)</td>
</tr>
<tr>
<td>Affirming</td>
<td>68.00* (23.95)</td>
<td>7.33 (1.26)</td>
<td>3.83* (1.02)</td>
<td>3.99 (1.23)</td>
</tr>
<tr>
<td>Corrective</td>
<td>68.06* (21.67)</td>
<td>7.20 (1.15)</td>
<td>3.75* (1.02)</td>
<td>3.60† (1.29)</td>
</tr>
</tbody>
</table>

*Significantly different from control group at p < .05.
†Different from control group at p < .09.
Table 14

*Study 2. Mean (SD) ratings of the probability that the defendant was the source of the evidence dependent on which article was read for low, moderate, and high NC participants.*

<table>
<thead>
<tr>
<th>NC group</th>
<th>Control</th>
<th>Complex</th>
<th>Simple</th>
<th>Affirming</th>
<th>Corrective</th>
</tr>
</thead>
<tbody>
<tr>
<td>High NC</td>
<td>82.38</td>
<td>69.59</td>
<td>77.60</td>
<td>69.36</td>
<td>62.22*</td>
</tr>
<tr>
<td></td>
<td>(20.46)</td>
<td>(19.45)</td>
<td>(20.70)</td>
<td>(23.43)</td>
<td>(22.95)</td>
</tr>
<tr>
<td>Moderate NC</td>
<td>80.40</td>
<td>75.52</td>
<td>65.77</td>
<td>75.22</td>
<td>67.20</td>
</tr>
<tr>
<td></td>
<td>(16.36)</td>
<td>(17.70)</td>
<td>(28.07)</td>
<td>(23.36)</td>
<td>(22.02)</td>
</tr>
<tr>
<td>Low NC</td>
<td>79.72</td>
<td>64.41</td>
<td>79.64</td>
<td>59.24*</td>
<td>76.05</td>
</tr>
<tr>
<td></td>
<td>(16.83)</td>
<td>(25.54)</td>
<td>(19.76)</td>
<td>(26.46)</td>
<td>(18.00)</td>
</tr>
</tbody>
</table>

*Significantly different from control group at $p < .05$.**
### Table 15

**Study 2. Mean (SD) ratings of perceived understanding of the expert’s testimony and his forensic technique (1-9) for low, moderate, and high NC participants.**

<table>
<thead>
<tr>
<th>NC Group</th>
<th>Expert Understanding</th>
<th>Evidence Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>High NC</td>
<td>7.99&lt;sup&gt;a&lt;/sup&gt; (1.26)</td>
<td>7.22&lt;sup&gt;a&lt;/sup&gt; (1.83)</td>
</tr>
<tr>
<td>Moderate NC</td>
<td>7.38&lt;sup&gt;b&lt;/sup&gt; (1.40)</td>
<td>6.73&lt;sup&gt;a&lt;/sup&gt; (1.80)</td>
</tr>
<tr>
<td>Low NC</td>
<td>7.22&lt;sup&gt;b&lt;/sup&gt; (1.75)</td>
<td>6.63&lt;sup&gt;a&lt;/sup&gt; (1.95)</td>
</tr>
</tbody>
</table>

*Within each DV, means with different superscripts significantly different (p < .05).*
Table 16

Study 3. Validity-belief ratings for control participants only for each type of evidence and results of paired-sample t-tests for each evidence comparison and average knowledge ratings for each evidence type.

<table>
<thead>
<tr>
<th>Validity-belief</th>
<th>Fingerprint</th>
<th>Toolmark</th>
<th>Bitemark</th>
<th>Eyewitness</th>
<th>Confession</th>
<th>Polygraph</th>
<th>Toxicology</th>
<th>Arson</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 92.30</td>
<td>t (63) = 3</td>
<td>t (63) = 15.77</td>
<td>t (63) = 12.09</td>
<td>t (63) = 14.09</td>
<td>t (63) = 5.95</td>
<td>t (63) = 9.44</td>
<td>M = 55.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 14.22</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fingerprint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 87.34</td>
<td>t (63) = 26.84</td>
<td>t (63) = 31.41</td>
<td>t (63) = 31.44</td>
<td>t (63) = 31.32</td>
<td>t (63) = 31.32</td>
<td>t (63) = 31.32</td>
<td>M = 31.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 16.36</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toolmark</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 57.19</td>
<td>t (63) = 42.66</td>
<td>t (63) = 52.56</td>
<td>t (63) = 58.84</td>
<td>t (63) = 55.89</td>
<td>t (63) = 55.89</td>
<td>t (63) = 55.89</td>
<td>M = 55.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 24.58</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitemark</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 65.56</td>
<td>t (63) = 1.19</td>
<td>t (63) = 4.39</td>
<td>t (63) = 3.54</td>
<td>t (63) = 1.19</td>
<td>t (63) = 1.98</td>
<td>t (63) = 1.98</td>
<td>M = 23.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 29.27</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eyewitness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 36.13</td>
<td>t (63) = -3.83</td>
<td>t (63) = 4.39</td>
<td>t (63) = 3.54</td>
<td>t (63) = 1.19</td>
<td>t (63) = 1.98</td>
<td>t (63) = 1.98</td>
<td>M = 23.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 24.27</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confession</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 46.24</td>
<td>t (63) = -7.45</td>
<td>t (63) = 4.39</td>
<td>t (63) = 3.54</td>
<td>t (63) = 1.19</td>
<td>t (63) = 1.98</td>
<td>t (63) = 1.98</td>
<td>M = 23.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 27.95</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polygraph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 41.16</td>
<td>t (63) = -10.55</td>
<td>t (63) = 4.39</td>
<td>t (63) = 3.54</td>
<td>t (63) = 1.19</td>
<td>t (63) = 1.98</td>
<td>t (63) = 1.98</td>
<td>M = 23.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 26.98</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toxicology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 79.68</td>
<td>t (63) = 6.59</td>
<td>t (63) = 3.54</td>
<td>t (63) = 1.98</td>
<td>t (63) = 1.98</td>
<td>t (63) = 1.98</td>
<td>t (63) = 1.98</td>
<td>M = 23.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 20.23</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = 63.30</td>
<td>t (63) = 5.88</td>
<td>t (63) = 4.39</td>
<td>t (63) = 3.54</td>
<td>t (63) = 1.19</td>
<td>t (63) = 1.98</td>
<td>t (63) = 1.98</td>
<td>M = 23.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 25.19</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
<td>155</td>
</tr>
</tbody>
</table>
Table 17

*Study 3. Mean (SD) validity-belief ratings in the different article conditions for bitemark and other evidence types.*

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Complex</th>
<th>Simple</th>
<th>Story</th>
<th>Story-and-Fact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitemark</td>
<td>65.56</td>
<td>25.73*</td>
<td>26.63*</td>
<td>54.08†</td>
<td>24.35*</td>
</tr>
<tr>
<td></td>
<td>(29.27)</td>
<td>(23.22)</td>
<td>(21.10)</td>
<td>(24.95)</td>
<td>(19.21)</td>
</tr>
<tr>
<td>DNA</td>
<td>92.30</td>
<td>89.67</td>
<td>88.36</td>
<td>93.26</td>
<td>93.50</td>
</tr>
<tr>
<td></td>
<td>(14.22)</td>
<td>(15.65)</td>
<td>(21.19)</td>
<td>(8.97)</td>
<td>(12.14)</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>87.34</td>
<td>82.35</td>
<td>76.84*</td>
<td>81.78</td>
<td>85.14</td>
</tr>
<tr>
<td></td>
<td>(16.36)</td>
<td>(20.82)</td>
<td>(23.08)</td>
<td>(19.34)</td>
<td>(15.61)</td>
</tr>
<tr>
<td>Toolmark</td>
<td>57.19</td>
<td>40.55*</td>
<td>41.56*</td>
<td>53.21</td>
<td>50.64</td>
</tr>
<tr>
<td></td>
<td>(24.58)</td>
<td>(23.10)</td>
<td>(24.31)</td>
<td>(23.81)</td>
<td>(25.88)</td>
</tr>
<tr>
<td>Eyewitness</td>
<td>36.13</td>
<td>35.06</td>
<td>34.84</td>
<td>39.72</td>
<td>37.26</td>
</tr>
<tr>
<td></td>
<td>(24.37)</td>
<td>(20.45)</td>
<td>(25.18)</td>
<td>(22.97)</td>
<td>(19.16)</td>
</tr>
<tr>
<td>Confession</td>
<td>46.24</td>
<td>46.84</td>
<td>44.62</td>
<td>50.03</td>
<td>49.34</td>
</tr>
<tr>
<td></td>
<td>(27.95)</td>
<td>(26.61)</td>
<td>(28.39)</td>
<td>(25.98)</td>
<td>(25.11)</td>
</tr>
<tr>
<td>Polygraph</td>
<td>41.16</td>
<td>35.11</td>
<td>32.71†</td>
<td>48.36†</td>
<td>39.89</td>
</tr>
<tr>
<td></td>
<td>(26.98)</td>
<td>(25.07)</td>
<td>(22.87)</td>
<td>(25.46)</td>
<td>(26.17)</td>
</tr>
<tr>
<td>Toxicology</td>
<td>79.68</td>
<td>71.78</td>
<td>75.14</td>
<td>81.52</td>
<td>77.81</td>
</tr>
<tr>
<td></td>
<td>(20.23)</td>
<td>(26.11)</td>
<td>(20.78)</td>
<td>(17.46)</td>
<td>(19.61)</td>
</tr>
<tr>
<td>Arson</td>
<td>63.30</td>
<td>53.70†</td>
<td>56.80</td>
<td>70.92†</td>
<td>61.82</td>
</tr>
<tr>
<td></td>
<td>(25.02)</td>
<td>(24.15)</td>
<td>(25.27)</td>
<td>(43.04)</td>
<td>(21.65)</td>
</tr>
</tbody>
</table>

*Significantly different from control group at $p < .05$.  
†Different from control group at $p < .06$.  
‡Within non-bitemark evidence types, article means significantly differ from each other ($p < .05$).
Table 18

Study 3. *Mean (SD) ratings of the probability that the defendant was the source of the bitemark or toolmark evidence dependent on which article was read.*

<table>
<thead>
<tr>
<th>Article Type</th>
<th>Bitemark</th>
<th>Toolmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>80.33 (25.04)</td>
<td>71.09 (27.32)</td>
</tr>
<tr>
<td>Complex</td>
<td>65.23 (20.96)</td>
<td>65.68 (32.07)</td>
</tr>
<tr>
<td>Simple</td>
<td>67.13 (24.80)</td>
<td>77.52 (21.22)</td>
</tr>
<tr>
<td>Story</td>
<td>70.66 (22.49)</td>
<td>71.77 (26.26)</td>
</tr>
<tr>
<td>Story-and-fact</td>
<td>57.31* (26.53)</td>
<td>78.86 (21.01)</td>
</tr>
</tbody>
</table>

*Significantly different from control group at $p < .05$. 
Table 19

*Study 3. Mean (SD) ratings of the credibility of the expert (1-9), and perceived understanding of the expert’s testimony and his forensic technique (1-9) dependent on which article was read.*

<table>
<thead>
<tr>
<th>Article type</th>
<th>Expert Credibility</th>
<th>Understanding Expert</th>
<th>Understanding Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>7.61 (1.34)</td>
<td>7.81 (1.36)</td>
<td>7.31 (1.38)</td>
</tr>
<tr>
<td>Complex</td>
<td>6.72* (1.39)</td>
<td>6.87* (1.68)</td>
<td>6.32* (1.92)</td>
</tr>
<tr>
<td>Simple</td>
<td>7.14 (1.49)</td>
<td>7.51 (1.49)</td>
<td>6.26* (1.78)</td>
</tr>
<tr>
<td>Story</td>
<td>7.34 (1.33)</td>
<td>7.51 (1.50)</td>
<td>6.66 (1.95)</td>
</tr>
<tr>
<td>Story-and-fact</td>
<td>6.95* (1.27)</td>
<td>7.44 (1.44)</td>
<td>6.75 (1.57)</td>
</tr>
</tbody>
</table>

*Significantly different from control group at p < .05.
Table 20

*Study 3. Mean (SD) ratings of the validity of the evidence (1-6) overall and in bitemark and toolmark conditions dependent on which article was read.*

<table>
<thead>
<tr>
<th>Article type</th>
<th>Overall</th>
<th>Bitemark</th>
<th>Toolmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>4.79 (0.93)</td>
<td>4.73 (0.95)</td>
<td>4.84 (0.93)</td>
</tr>
<tr>
<td>Complex</td>
<td>3.88* (1.06)</td>
<td>3.51* (1.06)</td>
<td>4.21† (0.96)</td>
</tr>
<tr>
<td>Simple</td>
<td>3.89* (1.14)</td>
<td>3.48* (1.12)</td>
<td>4.33 (1.00)</td>
</tr>
<tr>
<td>Story</td>
<td>4.50 (0.87)</td>
<td>4.41 (0.89)</td>
<td>4.62 (0.84)</td>
</tr>
<tr>
<td>Story-and-fact</td>
<td>4.10* (1.09)</td>
<td>3.41* (1.05)</td>
<td>4.70 (0.70)</td>
</tr>
</tbody>
</table>

*Significantly different from control group at p < .05.
†Different from control group at p < .06.
Table 21

*Study 3. Mean (SD) ratings of perceived understanding of the expert’s testimony and his forensic technique (1-9) for low, moderate, and high NC participants.*

<table>
<thead>
<tr>
<th>NC Group</th>
<th>Expert Understanding</th>
<th>Evidence Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>High NC</td>
<td>8.05\textsuperscript{a} (1.24)</td>
<td>7.15\textsuperscript{a} (1.56)</td>
</tr>
<tr>
<td>Moderate NC</td>
<td>7.31\textsuperscript{b} (1.40)</td>
<td>6.60\textsuperscript{b} (1.59)</td>
</tr>
<tr>
<td>Low NC</td>
<td>7.06\textsuperscript{b} (1.65)</td>
<td>6.44\textsuperscript{b} (1.94)</td>
</tr>
</tbody>
</table>

\*Within each DV, means with different superscripts significantly different ($p < .05$).
Appendix A: Informed Consent Study One

You are invited to participate in a research study entitled “Juror Decision Making in a Homicide Case.” The purpose of this research is to examine how jurors respond to different types of testimony in a trial regarding a violent criminal act. We plan to enroll approximately 900 participants into this study. If you decide to participate, you will be asked to read a trial summary and then answer a number of questions. Participation should take about one hour in a single session.

The foreseeable risks of participation in this study are minimal. Although you will read about a murder, the content will be similar to that presented every day in the news and popular media. If you anticipate that reading about a violent crime will be distressing to you, please alert the researcher. In order to minimize these risks we will discontinue the study immediately if you feel any discomfort during the study. The possible benefits to you are will see how research is conducted and will learn more about the topic under study. The potential benefits to society are that we will learn more about how jurors respond to different types of testimony in a criminal trial.

Please check to confirm that you have read and understand the potential risks and benefits of your participation:

Your participation in this study is completely voluntary. You have a right to refuse to participate without consequences. If you decide not to participate your decision will not affect your relationship with John Jay College or with the researchers conducting this or future studies.

If you decide to participate you may discontinue participation at any time. You may refuse to answer any specific questions or refuse to engage in any task at any time during the study. Withdrawal or refusing to answer specific questions or engage in specific tasks will not result in any consequences to you and will not affect your relationship with John Jay College or with the researchers conducting this or future studies.

Please check to confirm that you have read and understand that participation is voluntary:

Information gathered from you will be kept strictly confidential. You will not be identified individually in any way as a result of your participation in this research. Your informed consent is being collected on a separate website from your responses to the study itself; thus, your name will be kept totally separate from your data. Additionally, any printed data sheets will be kept in the locked file cabinets in the offices of Dr. Maureen O’Connor and/or Victoria Lawson. After a period of five years, any printed data sheets will be destroyed and the file containing informed consent data and the file containing your original responses will be deleted.

Checking the box below means that you have read this consent form, that you fully understand the nature and consequences of participation and that you have had all questions regarding participation in this study answered satisfactorily. If you have further questions about this research please feel free to contact the Principal Investigator, Maureen O’Connor at moconnor@jjay.cuny.edu or her research assistant, Victoria Lawson at vlawson@jjay.cuny.edu.

If you have any questions regarding your rights as a research participant please feel free to contact the John Jay Institutional Review Board Office at jj-irb@jjay.cuny.edu, or (212) 237-8961.

I have read this consent form carefully and I give my consent to participate in this study: ☐
Appendix B: Study One Post-trial Questionnaire

Manipulation Checks

What was the defendant charged with?

a. Arson  
b. Murder  
c. Burglary  
d. Kidnapping

Where were the defendant and victim seen together?

a. Diner  
b. Arcade  
c. Bar  
d. Coffee Shop

Did the defendant testify?

No  
Yes

Did an expert witness testify about whether or not physical evidence connected the defendant to the crime? (Note: This does not include the detective.)

No  
Yes

Experimental participants only:

If yes, what type of expert witness testified?

a. DNA expert  
b. Fingerprint expert  
c. Bitemark expert  
d. Toolmark expert

Based on his forensic analysis, did the expert link the defendant, Michael Smith, to the rubber tubing used in the crime?

No  
Yes

In a few sentences, please describe the most important points of the expert’s testimony:

Verdict and Perceptions of Guilt Questions (all participants)

If you were a juror in this case, how would you vote on the verdict?

Not Guilty  
Guilty
How confident are you in your verdict choice?

1  2  3  4  5  6  7
Not at all confident  Very confident

Please give an estimate of how likely you think it is that the defendant murdered the victim on a scale from 0 (definitely did not murder the victim) to 100 (definitely murdered the victim): ____

Experimental participants only:

Please give an estimate of how likely you think it is that the defendant was the source of the saliva on the rubber tubing on a scale from 0 (definitely not the source) to 100 (definitely the source): ____

Perceptions of the Expert and Evidence Questions

Witness credibility questions (used for all witnesses including expert):

Please rate the extent to which you felt that _____ was:

<table>
<thead>
<tr>
<th>Believable</th>
<th>1 (Not at all)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7 (Very)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dishonest</td>
<td>1 (Not at all)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7 (Very)</td>
</tr>
<tr>
<td>Convincing</td>
<td>1 (Not at all)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7 (Very)</td>
</tr>
<tr>
<td>Incompetent</td>
<td>1 (Not at all)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7 (Very)</td>
</tr>
<tr>
<td>Trustworthy</td>
<td>1 (Not at all)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7 (Very)</td>
</tr>
</tbody>
</table>

How well did you understand the testimony of ____:

1  2  3  4  5  6  7
(Not at all well) (Very well)
Only for expert (not all witnesses):

How well did you understand the forensic technique used by the expert witness Richard Williams?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Not at all well)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Very well)</td>
</tr>
</tbody>
</table>

Evidence quality questions (experimental participants only)

Please indicate the extent to which you agree with the following statements about the forensic identification technique used by expert witness Richard Williams.

The forensic identification technique used by expert witness Williams was based on good scientific principles.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disagree strongly</td>
<td>Disagree</td>
<td>Somewhat disagree</td>
<td>Somewhat agree</td>
<td>Agree</td>
<td>Agree strongly</td>
</tr>
</tbody>
</table>

The forensic identification technique used by expert witness Williams was not reliable.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disagree strongly</td>
<td>Disagree</td>
<td>Somewhat disagree</td>
<td>Somewhat agree</td>
<td>Agree</td>
<td>Agree strongly</td>
</tr>
</tbody>
</table>

The forensic identification technique used by expert witness Williams was not appropriate in this case.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disagree strongly</td>
<td>Disagree</td>
<td>Somewhat disagree</td>
<td>Somewhat agree</td>
<td>Agree</td>
<td>Agree strongly</td>
</tr>
</tbody>
</table>

The methods used by expert witness Williams to test the evidence were sound.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disagree strongly</td>
<td>Disagree</td>
<td>Somewhat disagree</td>
<td>Somewhat agree</td>
<td>Agree</td>
<td>Agree strongly</td>
</tr>
</tbody>
</table>

The forensic identification technique used by expert witness Williams was helpful in reaching my verdict.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disagree strongly</td>
<td>Disagree</td>
<td>Somewhat disagree</td>
<td>Somewhat agree</td>
<td>Agree</td>
<td>Agree strongly</td>
</tr>
</tbody>
</table>
The forensic identification technique used by expert witness Williams was not scientifically valid.

1 Disagree strongly
2 Disagree
3 Somewhat disagree
4 Somewhat agree
5 Agree
6 Agree strongly

The forensic identification technique used by expert witness Williams did not contribute strongly to my verdict.

1 Disagree strongly
2 Disagree
3 Somewhat disagree
4 Somewhat agree
5 Agree
6 Agree strongly
Appendix C: Informed Consent Studies Two and Three

**Introduction/Purpose:** You are invited to participate in a research study. The study is conducted under the direction of Victoria Lawson and Deryn Strange, John Jay College of Criminal Justice. The purpose of this study is to examine what makes criminal justice articles interesting and how to make them more compelling; and to examine how jurors respond to different types of testimony in criminal trials. The results of this study may aid in determining the best way to present criminal justice information in the news media and to examine how jurors respond to different types of testimony in criminal trials.

**Procedures:** Approximately 400 individuals are expected to participate in this study. Each participant will read some articles and then answer some questions about them; next, on a different website, you will read the transcript of a homicide trial and then answer some questions about it. The time commitment of each participant is expected to be approximately 1-1.5 hours and all data collection will be completed online in one session across two websites.

**Possible Discomforts and Risks:** Your participation in this study may involve some slight discomfort, as you will be reading about topics related to the criminal justice system and—on the second website—you will be reading about a violent criminal act; however, the information provided is of a type commonly presented in the news and popular media. If you are at all upset as a result of this study, a searchable database of mental health services throughout the country is available at the following address: [http://store.samhsa.gov/mhlocator](http://store.samhsa.gov/mhlocator)

**Benefits:** The direct benefits to you are that you will be compensated for your participation. Indirect benefits are that you may learn more about how research is conducted and about the topic under study; you may also learn new information from the articles, and may learn about the evidence presented in criminal trials.

**Voluntary Participation:** Your participation in this study is voluntary, and you may decide not to participate without prejudice, penalty, or loss of benefits to which you are otherwise entitled. If you decide to leave the study, you may simply close this website; no further action is needed on your part.

**Financial Considerations:** For your participation you will receive $2. As soon as we have confirmed your participation through your anonymous ID number, we will approve your HIT on Mechanical Turk, releasing the funds.

**Confidentiality:** The data obtained from you will be collected via PsychSurveys. The collected data will be accessible to the researchers named above, Victoria Lawson and Dr. Deryn Strange. In order to ensure your confidentiality, the researcher not be collecting any identifiable information from you. The collected data will be stored on the password-protected computers of the researchers named above, Victoria Lawson and Deryn Strange.

**Contact Questions/Persons:** If you have any questions about the research now or in the future, you should contact the Principal Investigator, Victoria Lawson, 646-557-4853,
If you have any questions concerning your rights as a participant in this study, you may contact Carina Quintian, HRPP Coordinator, 212-237-8961, jj-irb@jjay.cuny.edu.

**Statement of Consent:**

“I have read the above description of this research and I understand it. I have been informed of the risks and benefits involved, and all my questions have been answered to my satisfaction. Furthermore, I have been assured that any future questions that I may have will also be answered by the principal investigator of the research study. I voluntarily agree to participate in this study.

By checking the box below I have not waived any of my legal rights to which I would otherwise be entitled.

I may print a copy of this statement for my records.”

---

**I have read this form and understand my rights as a research participant:** ☐

**I give my consent to participate in this study:** ☐
Appendix D: Complex Article

‘Badly fragmented’ forensic science system needs total overhaul: Study

WASHINGTON - A congressionally mandated report from the National Academy of Sciences found serious deficiencies in the nation's forensic science system and calls for major reforms and new research. According to the report, rigorous and mandatory certification programs for forensic scientists are currently lacking, as are strong standards and protocols for analyzing and reporting on evidence. And there is a dearth of peer-reviewed, published studies establishing the scientific bases and reliability of many forensic methods. Moreover, many forensic science labs are underfunded, understaffed, and have no effective oversight.

Forensic evidence is often offered in criminal prosecutions and civil litigation to support conclusions about individualization -- in other words, to "match" a piece of evidence to a particular person, weapon, or other source. But with the exception of nuclear DNA analysis, the report says, no forensic method has been rigorously shown able to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source. Other forensic disciplines have important roles, but many need substantial research to validate basic premises and techniques, assess limitations, and discern the sources and magnitude of error, according to the report.

The report explicitly recognizes that there's a world of difference between forensic science as an investigatory tool or method and proof of innocence or guilt in court. One example of this type of forensic work cited the report is forensic odontology. Forensic odontology, the application of the science of dentistry to the field of law, includes several areas of focus: the interpretation of unknown remains, bite mark comparison, the interpretation of oral injury, and dental malpractice. Bite mark comparison is often used in criminal prosecutions and is the most controversial of the four areas just mentioned and was criticized heavily in the report.

The report noted that no scientific studies support the assertion of forensic odontologists that bite marks can demonstrate sufficient details for positive identification and that no large population studies have been conducted. It further recognized that in numerous instances, experts diverge widely in their evaluations of the same bite mark evidence, which has led to questioning of the value and scientific objectivity of such evidence.

The report pointed to several basic problems inherent in bite mark analysis and interpretation. Specifically, it noted that the uniqueness of the human dentition has not been scientifically established. The ability of the dentition, if unique, to transfer a unique pattern to human skin and the ability of the skin to maintain that uniqueness has not been scientifically established. Further, the ability to analyze and interpret the scope or extent of distortion of bite mark patterns on human skin has not been demonstrated, and the effect of distortion on different comparison techniques is not fully understood and therefore has not been quantified. Finally, a standard for the type, quality, and number of individual characteristics required to indicate that a bite mark has reached a threshold of evidentiary value has not been established.
Appendix E: Simple Article

‘Badly fragmented’ forensic science system needs total overhaul: Study

WASHINGTON – According to a report released today by the National Academy of Sciences, major reforms are needed in the forensic sciences. Their study was requested by the United States Congress and its aim was to assess whether the different forensic sciences are actually valid and reliable. To answer that question, the authors considered whether there is good quality research to support both the techniques and the conclusions that practitioners draw after conducting their analyses. They also examined the qualifications required for practitioners. Overall, the report concluded that there is not enough research on the majority of the forensic sciences and therefore their techniques and conclusions cannot be considered valid. Moreover, the authors found that many of the sciences did not have strict standards or certification requirements to actually practice the science.

In fact, the report found that only nuclear DNA analysis can be reliably and accurately used to draw conclusions about who or what left behind a particular piece of evidence. The authors found that no other forensic science should be used to link any particular suspect to any particular piece of crime scene evidence. These other sciences can be used to show that someone did not leave a piece of evidence, but at present they cannot say who did.

Forensic odontology—the practice of dentistry as it is used in the legal field—was one of the forensic sciences studied in the report. There are different practices included in the area, but the most controversial is bite mark comparison. While bite mark comparisons are often used in criminal prosecutions, they were criticized heavily in the report.

The biggest criticism concerned the claim that bite mark analysts can actually identify a person as the source of any particular bite mark; the report concluded that bite mark comparison cannot actually be used to identify a bite mark’s source. Instead, the report found that there is not enough scientific evidence to show that people’s dentitions—the positions and characteristics of their teeth—are actually unique.

First, the authors noted that there has been no research showing how common it is for people’s teeth to make the same mark. That means that even if a person’s teeth are found to match a particular bite mark perfectly, there is no way to know how many other people’s teeth would match just as well. Second, even if future research could show that people’s teeth are unique, it is still not clear whether the lasting marks they make on skin are also unique or whether important characteristics are distorted or unable to be viewed. It is possible that the mark might not match particularly well to the teeth that left it, or that important identifiers might not be visible.

The report concluded that bite mark comparison should only be used to exclude suspects—say who could not have left a particular mark—but should never be used to say that a specific person’s teeth made any particular bite mark.
Appendix F: Affirming Article

‘Badly fragmented’ forensic science system needs total overhaul: Study

WASHINGTON – According to a report released today by the National Academy of Sciences, major reforms are needed in the forensic sciences. Their study was requested by the United States Congress. In explaining the need for the study, Congress recognized that these sciences are law enforcement’s best tool for linking a suspect to a crime and ensuring that he or she can be successfully prosecuted. They noted that there have been increasing concerns about the reliability of the forensic sciences, concerns which have led convictions to be overturned and may in future make it difficult for these sciences to be admitted in court at all. Thus, the study’s primary goal was to assess whether the different forensic sciences are actually valid and reliable and how to improve them so that they can be relied upon with confidence by law enforcement and the courts.

To answer these questions, the authors considered whether there is good quality research to support both the techniques and the conclusions that practitioners draw after conducting their analyses. They also examined the qualifications required for practitioners. Overall, the report concluded that there is not enough research on the majority of the forensic sciences and therefore their techniques and conclusions cannot be considered valid. Moreover, the authors found that many of the sciences did not have strict standards or certification requirements to actually practice the science, which could prevent practitioners from being recognized as experts in court.

In fact, the report found that only nuclear DNA analysis can be reliably and accurately used to draw conclusions about who or what left behind a particular piece of evidence. The authors found that no other forensic science should be used to link any particular suspect to any particular piece of crime scene evidence. These other sciences can be used to show that someone did not leave a piece of evidence, but at present they cannot say who did.

Forensic odontology—the practice of dentistry as it is used in the legal field—was one of the forensic sciences studied in the report. There are different practices included in the area, but the most controversial is bite mark comparison. While bite mark comparisons are often used in criminal prosecutions, they were criticized heavily in the report.

The biggest criticism concerned the claim that bite mark analysts can actually identify a person as the source of any particular bite mark; the report concluded that bite mark comparison cannot actually be used to identify a bite mark’s source. Instead, the report found that there is not enough scientific evidence to show that people’s dentitions—the positions and characteristics of their teeth—are actually unique.

First, the authors noted that there has been no research showing how common it is for people’s teeth to make the same mark. That means that even if a person’s teeth are found to match a particular bite mark perfectly, there is no way to know how many other people’s teeth would match just as well. Second, even if future research could show that people’s teeth are unique, it is still not clear whether the lasting marks they make on skin are also unique or whether important
characteristics are distorted or unable to be viewed. It is possible that the mark might not match particularly well to the teeth that left it, or that important identifiers might not be visible.

The report concluded that bite mark comparison should only be used to exclude suspects—say who could not have left a particular mark—but should never be used to say that a specific person’s teeth made any particular bite mark.
Appendix G: Corrective Article

‘Badly fragmented’ forensic science system needs total overhaul: Study

WASHINGTON – According to a report released today by the National Academy of Sciences, major reforms are needed in the forensic sciences. Their study was requested by the United States Congress and its aim was to assess whether the different forensic sciences are actually valid and reliable. To answer that question, the authors considered whether there is good quality research to support both the techniques and the conclusions that practitioners draw after conducting their analyses. They also examined the qualifications required for practitioners. Overall, the report concluded that there is not enough research on the majority of the forensic sciences and therefore their techniques and conclusions cannot be considered valid. Moreover, the authors found that many of the sciences did not have strict standards or certification requirements to actually practice the science.

In fact, the report found that only nuclear DNA analysis can be reliably and accurately used to draw conclusions about who or what left behind a particular piece of evidence. The authors found that no other forensic science should be used to link any particular suspect to any particular piece of crime scene evidence. These other sciences can be used to show that someone did not leave a piece of evidence, but at present they cannot say who did.

Forensic odontology—the practice of dentistry as it is used in the legal field—was one of the forensic sciences studied in the report. There are different practices included in the area, but the most controversial is bite mark comparison. While bite mark comparisons are often used in criminal prosecutions, they were criticized heavily in the report.

The biggest criticism concerned the claim that bite mark analysts can actually identify a person as the source of any particular bite mark; the report concluded that bite mark comparison cannot actually be used to identify a bite mark’s source. Instead, the report found that there is not enough scientific evidence to show that people’s dentitions—the positions and characteristics of their teeth—are actually unique.

First, the authors noted that there has been no research showing how common it is for people’s teeth to make the same mark. That means that even if a person’s teeth are found to match a particular bite mark perfectly, there is no way to know how many other people’s teeth would match just as well. Second, even if future research could show that people’s teeth are unique, it is still not clear whether the lasting marks they make on skin are also unique or whether important characteristics are distorted or unable to be viewed. It is possible that the mark might not match particularly well to the teeth that left it, or that important identifiers might not be visible.

The report concluded that bite mark comparison should only be used to exclude suspects—say who could not have left a particular mark—but should never be used to say that a specific person’s teeth made any particular bite mark. They noted that their findings have implications for how bite mark evidence should be viewed if presented at trial. If an expert testifies that a suspect has been excluded—that he or she did not leave a mark—then jurors can be fairly certain that the suspect did not, in fact, leave the mark. If, on the other hand, an expert testifies that a suspect’s
teeth match a mark, jurors should not conclude that the suspect has left the mark, but that he or she is one of an unknown number of people whose teeth match certain characteristics as identified by the practitioner. Thus a bite mark match in itself should not be considered sufficient evidence to support a verdict of guilt; jurors should convict only if there is additional convincing evidence.
Appendix H: Study Two Post-Target-Article Questions

How easy or difficult was it to understand the article you just read?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very easy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neither easy nor difficult</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very difficult</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How clearly were the findings of the report described in the article?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all clearly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very clearly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please briefly describe the most important points of the article and what you found most interesting:

All target articles:

What did the National Academy of Sciences’ report conclude regarding the certification requirements of forensic scientists?

a. Rigorous certification requirements exist for most disciplines
b. Many of the disciplines do not have strict certification requirements or standards
c. Certification requirements are not recommended because they are a poor indicator of an expert’s qualifications
d. Standardized certification requirements exist across disciplines

What is a major problem with drawing conclusions from bite mark comparisons?

a. There are no problems with drawing conclusions from bite mark comparisons
b. There are too few qualified forensic odontologists
c. It is not known whether human teeth are unique
d. Most of the research conducted on bite mark analysis is outdated

Which forensic science(s) did the report say can be used to determine whether a suspect is the source of crime scene evidence?

a. Nuclear DNA and fingerprint analysis
b. DNA, toolmark, bite mark, fingerprint, and arson analysis
c. Most of the forensic sciences
d. Nuclear DNA only
Complex article only:

The report criticized bite mark analysis for which of the following reasons?

a. Forensic odontologists charge more for their services than experts in other disciplines
b. No scientific studies support the claim that bite marks can demonstrate enough detail for a positive identification
c. Too often experts are conservative in their judgments which can lead them to say that too few bite marks match
d. The report had no criticisms of bite mark analysis

Simple article only:

Which of the following is a reason that the report gave for saying that there is not enough scientific evidence to show that people’s dentitions are unique?

a. Although most research suggests that they are unique, no meta-analyses have been conducted on the topic
b. It is not known how many people’s teeth could match a particular bite mark
c. Most of the research to date fails to account for modern dentistry
d. Experts might be looking at the wrong characteristics entirely

Affirming article only:

What is one reason that Congress said that a study on the forensic sciences was necessary?

a. To show that the forensic sciences are not reliable and should be eliminated
b. To ensure that the sciences are valid so that suspects can be successfully prosecuted
c. To prove critics of the forensic sciences wrong by showing that they are 100% reliable in all circumstances
d. To ensure that scientists know how to perform the analyses used in their field and do not make errors

Corrective article only:

According to the authors of the report, what can jurors conclude from bite mark testimony?

a. If the expert testifies that the suspect’s teeth match a bite mark, jurors should conclude that the suspect definitely made the mark
b. If the expert testifies that the suspect’s teeth do not match a bite mark, jurors can be fairly certain that the suspect did not make the mark
c. If the expert testifies that a suspect’s teeth do not match a bite mark, jurors should ignore the expert’s testimony
d. If the expert testifies that a suspect’s teeth match a bite mark, jurors should conclude that the suspect committed the crime
Appendix I: Story Article

Charges dropped in Wisconsin case

MILWAUKEE – Milwaukee prosecutors announced Monday that they will not seek a new trial in the case of Robert Lee Stinson, who served more than 23 years in prison for a murder DNA proves he didn’t commit. He was released in January after his conviction was overturned, but Monday’s announcement makes his exoneration official.

Stinson was convicted in 1985 for the murder of a 63-year-old woman in Milwaukee. A neighbor passing through an alley on his way to work discovered her body in a vacant lot behind her home. She had been raped, stabbed, and beaten to death. Eight bite marks, inflicted prior to death, were also identified on the victim’s body.

After examining the body, dental scientist Dr. Lowell Thomas Johnson worked with a police sketch artist and determined that the bite marks on the body must have come from someone missing an upper front tooth. During the initial inquiry, two police investigators visited Stinson, who was a neighbor of the victim. While interviewing Stinson, the investigators told him a joke, and noticed that he was missing a front tooth when he laughed. Based on this observation, and his proximity to the crime scene, Stinson was arrested and charged with murder.

Stinson, 21 at the time of his arrest, proclaimed his innocence from the witness stand. Although he gave somewhat inconsistent accounts of his whereabouts at the time of the murder, the only evidence directly connecting Stinson to the crime was the bite mark testimony of two forensic odontologists. Dr. Johnson had had models taken of Stinson’s teeth and taken photos of his dentition and of the bite marks. He concluded that the bite marks “had to have been made by teeth identical” to Stinson’s, and that there was “no margin for error” in his conclusion. The State also called Dr. Raymond Rawson, the chairman of the Bite Mark Standards Committee of the American Board of Forensic Odontologists, who testified that the evidence in the case was “high quality” and “overwhelming.” After a three-day trial, Stinson was convicted of first-degree murder on the strength of the forensic testimony, and sentenced to life in prison.

In 2005, Stinson wrote to the Wisconsin Innocence Project at the University of Wisconsin Law School, asking them for help with his case. He told them he had never been made eligible for parole because he refused to say that he committed the murder. After taking his case, they sought DNA testing on areas of the victim’s sweater that had tested positive for saliva, suggesting that these were the places where she had been bitten by her assailant. The testing revealed a male DNA profile, but that profile did not match Stinson, proving another person bit the victim. The next step will be to enter the profile into the state DNA database to see if a match can be found.

The DNA evidence was enough to persuade Circuit Judge Patricia McMahon to vacate Stinson's conviction. After ordering his release in January, she gave the state six months to decide whether or not to retry him. After undertaking their own investigation, they decided not to retry Stinson and officially announced that they would drop the charges in court on Monday. Their announcement officially exonerates Stinson and makes him eligible for compensation from the state.
Appendix J: Story-and-Fact Article

Charges dropped in Wisconsin case

MILWAUKEE – Milwaukee prosecutors announced Monday that they will not seek a new trial in the case of Robert Lee Stinson, who served more than 23 years in prison for a murder DNA proves he didn’t commit. He was released in January after his conviction was overturned, but Monday’s announcement makes his exoneration official.

Stinson was convicted in 1985 for the murder of a 63-year-old woman in Milwaukee. A neighbor passing through an alley on his way to work discovered her body in a vacant lot behind her home. She had been raped, stabbed, and beaten to death. Eight bite marks, inflicted prior to death, were also identified on the victim’s body.

After examining the body, dental scientist Dr. Lowell Thomas Johnson worked with a police sketch artist and determined that the bite marks on the body must have come from someone missing an upper front tooth. During the initial inquiry, two police investigators visited Stinson, who was a neighbor of the victim. While interviewing Stinson, the investigators told him a joke, and noticed that he was missing a front tooth when he laughed. Based on this observation, and his proximity to the crime scene, Stinson was arrested and charged with murder.

Stinson, 21 at the time of his arrest, proclaimed his innocence from the witness stand. Although he gave somewhat inconsistent accounts of his whereabouts at the time of the murder, the only evidence directly connecting Stinson to the crime was the bite mark testimony of two forensic odontologists. Dr. Johnson had had models taken of Stinson’s teeth and taken photos of his dentition and of the bite marks. He concluded that the bite marks “had to have been made by teeth identical” to Stinson’s, and that there was “no margin for error” in his conclusion. The State also called Dr. Raymond Rawson, the chairman of the Bite Mark Standards Committee of the American Board of Forensic Odontologists, who testified that the evidence in the case was “high quality” and “overwhelming.” After a three-day trial, Stinson was convicted of first-degree murder on the strength of the forensic testimony, and sentenced to life in prison.

In 2005, Stinson wrote to the Wisconsin Innocence Project at the University of Wisconsin Law School, asking them for help with his case. He told them he had never been made eligible for parole because he refused to say that he committed the murder. After taking his case, they sought DNA testing on areas of the victim’s sweater that had tested positive for saliva, suggesting that these were the places where she had been bitten by her assailant. The testing revealed a male DNA profile, but that profile did not match Stinson, proving another person bit the victim. The next step will be to enter the profile into the state DNA database to see if a match can be found.

Forensic odontology recently came under fire in the National Academy of Science’s Report on the forensic sciences which found major problems with the nation’s forensic science system. Their biggest criticism concerned the claim that bite mark analysts can actually identify a person as the source of any particular bite mark; the report concluded that bite mark comparison cannot actually be used to identify a bite mark’s source. Instead, the report found that there is not enough scientific evidence to show that people’s dentitions—the positions and characteristics of their teeth—are actually unique.
First, the authors noted that there has been no research showing how common it is for people’s teeth to make the same mark. That means that even if a person’s teeth are found to match a particular bite mark perfectly, there is no way to know how many other people’s teeth would match just as well. Second, even if future research could show that people’s teeth are unique, it is still not clear whether the lasting marks they make on skin are also unique or whether important characteristics are distorted or unable to be viewed. It is possible that the mark might not match particularly well to the teeth that left it, or that important identifiers might not be visible.

The report concluded that bite mark comparison should only be used to exclude suspects—say who could not have left a particular mark—but should never be used to say that a specific person’s teeth made any particular bite mark.
Appendix K: Study Three Story-Based Article Questions

Why was Robert Lee Stinson initially arrested?

a. Because an eyewitness saw him with the victim the night she was killed
b. Because he was missing a tooth and was a neighbor of the victim
c. Because he confessed to murdering the victim
d. Because of his close relationship with the victim

How many forensic odontologists (bite mark experts) testified at Stinson’s trial?

a. Three
b. Five
c. Two
d. None

What did the Wisconsin Innocence Project seek DNA testing on?

a. Swabs from the victim’s rape kit
b. A bloodstain on Stinson’s shirt
c. A bloodstain on the victim’s pants
d. Saliva on the victim’s sweater

Story article only:

Why wasn't Stinson made eligible for parole?

a. Because of a misunderstanding with the warden
b. Because he would not say that he committed the murder
c. Because the victim's family contested his release
d. Because he was involved in an altercation while in prison

Story-and-fact article only:

Which of the following is a reason that the National Academy of Sciences’ report gave for saying that there is not enough scientific evidence to show that people’s dentitions are unique?

a. Although most research suggests that they are unique, no meta-analyses have been conducted on the topic
b. It is not known how many people’s teeth could match a particular bite mark
c. Most of the research to date fails to account for modern dentistry
d. Experts might be looking at the wrong characteristics entirely
References


doi:10.1037//1076-8971.6.3.677


