UNIVERSITY OF CALIFORNIA,
IRVINE

The Impact of Narrative Consistency on Jurors' Utilization of DNA Evidence

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF ARTS

in Social Ecology

by

Lauren Ashley Reiser

Thesis Committee:
Assistant Professor Nicholas Scurich, Chair
Professor Linda J. Levine
Assistant Professor Paul K. Piff

2015

TABLE OF CONTENTS

ACKNOWLEDGMENTS

First and foremost, I would like to thank my committee chair, Dr. Nicholas Scurich, for his guidance and expertise during the completion of this thesis. Without his guidance and help this thesis would not have been possible. I would also like to express my gratitude to my committee members, Dr. Linda Levine and Dr. Paul Piff, for their invaluable advice and feedback on the paper itself.

ABSTRACT OF THE THESIS

The Impact of Narrative Consistency on Jurors' Utilization of DNA Evidence

By

Lauren Ashley Reiser

Master of Arts in Social Ecology

University of California, Irvine, 2015

Professor Nicholas Scurich, Chair

The Story Model of Juror Decision Making states that as jurors are exposed to new pieces of evidence, they continually integrate evidence into a "story" about what happened in the case; this process includes evaluating contradictory testimony and discounting evidence that does not fit within the juror's narrative about the case. Existing research has neglected to test how forensic DNA evidence is incorporated into jurors' narratives, especially if the DNA is inconsistent with the non-forensic evidence. The lack of emphasis on forensic evidence should be addressed given the perception of infallibility that surrounds DNA. Study 1 manipulated non-forensic evidence strength and whether there was a DNA match to test how jurors integrated DNA evidence into their narrative interpretation of the case. Results indicated that utilization of forensic testimony depended on the non-forensic evidence strength and reliance on non-forensic evidence depended on whether there was a DNA match. To expand the results from Study 1 and incorporate the possibility of error in DNA testing, Study 2 manipulated whether a DNA match did or did not contain laboratory error rate evidence and the strength of the non-forensic evidence to examine how jurors integrate error rates into their narrative about a case. Results showed that jurors were not sensitive to the laboratory error rates manipulation. Together, these studies indicate that jurors are not sensitive to the possibility of erroneous DNA results, but integrate DNA

identifications into their narrative; this suggests that bias might be integrated into their narratives that facilitate false convictions.

INTRODUCTION

In 1991, an eleven year old child was brutally raped and sodomized near her neighborhood swimming pool (Thompson, Taroni, & Aitken, 2003). Lacking much evidence beyond a vague description of the assailant and a semen sample, detectives began to focus the investigation on a local resident, Timothy Durham, who had a criminal record for firearm and parole violations. Durham was arrested and tried for rape. The prosecution presented the eyewitness' tentative identification and a DNA test that showed Durham's genotype matched that of the perpetrator's semen. Durham was convicted and sentenced to life in prison despite the testimony of 11 different alibi witnesses and time stamped bank statements proving Durham was in a different state when the rape occurred. Over three years after Durham's conviction, re-testing of the DNA from the crime scene showed that Durham's DNA did not, in fact, have the same genotype as the perpetrator, it differed on specific alleles, and matched DNA of a convicted serial rapist. The original DNA test results that made Durham look guilty had been an error and he was later released from prison (Thompson et al., 2003).

While DNA has been heralded as infallible, data from the Innocence Project indicate that forensic science errors are the second largest factor in wrongful convictions in America and errors have occurred in 63% of cases where individuals have later been exonerated (Saks & Koehler, 2005). Despite the possibility of error, DNA is considered extremely reliable and compelling (Lieberman, Carrell, Miethe, & Krauss, 2008), but, occasionally, DNA identification evidence and other case evidence are incongruent. It is important to evaluate how jurors respond to inconsistencies between forensic and non-forensic evidence. Therefore, the current study will investigate the influence of forensic DNA evidence in conjunction with non-forensic evidence to

determine how much weight jurors give to both types of evidence and how forensic and non-forensic evidence interact.

In the present research, we investigated how mock jurors utilize DNA evidence and non-forensic evidence in their evaluations of a legal case. Existing research suggests that jurors rely heavily on DNA evidence (Lieberman, et al., 2008), and interpret case facts by building a narrative interpretation of the evidence (Pennington & Hastie, 1988), but little research has incorporated forensic evidence into the narrative of the case. It is especially important to research this topic when the forensic and non-forensic evidence are inconsistent because of the risk that the compelling nature of reported DNA matches may transform how non-forensic evidence as a whole is interpreted. To test this, the following experiments assessed how mock jurors evaluate incompatible forensic and non-forensic evidence. Additionally, this research addresses whether jurors evaluate DNA evidence independently or whether forensic evidence becomes interrelated with non-forensic evidence presented during a trial.

DNA Testing and Errors

Forensic DNA evidence is an increasingly prominent tool in the criminal justice system (Lieberman et al., 2008). With few exceptions (e.g., identical twins), individuals have semi-unique genetic profiles, and advancements in scientific testing allow forensic scientists to compare DNA left at a crime scene with the DNA of a suspect. DNA matches are extremely accurate, however erroneous DNA test results can occur through sample contamination, mislabeling samples, deliberate human error, and accidental human error (Thompson, 2013).

Despite the possibility of forensic testing errors, testimony about DNA identification evidence is commonly allowed in court (Thompson, 2013). Courts address the possibility of error by prohibiting experts from making broad conclusions. For example, when experts testify

about DNA identifications in court, they are not allowed to state that the defendant's DNA matches the perpetrator's DNA because this is not technically accurate; forensic technicians analyze whether a DNA sample from a suspect matches DNA collected from the crime scene. There are two types of information that experts should include in their testimony about the reliability of DNA identification evidence that can indicate whether an error has occurred and a reported "match" result might not be a true match.

The first is the *random match probability* (RMP), which states the likelihood of a match occurring due to similar genetic profiles (Koehler, Chia, & Lindsey, 1995). When DNA tests are performed, the entire genetic sequence of Source A is not compared to the entire genetic sequence of Source B; instead, a *portion* of the genetic sequence from both genetic sources is compared (Thompson, 2013). While quite rare, this can yield a determination that the two samples match, when in fact the two samples did not come from one individual but rather two different people who, coincidentally, share a similar genetic composition. As a result, it is possible that a match could be falsely identified between a perpetrator and an unrelated individual's DNA. This coincidental profile match is rare, so the RMP between two unrelated people is often extremely low, for instance 1-in-200 trillion (Thompson, 2013).

The second distinct piece of information that DNA experts can testify about is the *laboratory error rate*—the likelihood of a laboratory DNA test returning a false positive or a false negative result (Koehler, 1996). A false positive test result occurs when a DNA match is reported, but the reported match is not a true match. In other words, an error has occurred at some point in the DNA analysis procedure, indicating that crime scene DNA matches a suspect when, in reality, the suspect is not the actual perpetrator. For example, if DNA from a crime scene is accidentally cross-contaminated with DNA taken from an innocent suspect, the transfer

of genetic materials will cause the test results to indicate that the suspect's DNA was in the sample collected from the crime scene even though he was not actually present (Thompson, 2013). This type of error can contribute to wrongful convictions because a defendant is not the source of the genetic material, whereas the test may indicate otherwise. A false negative test result occurs when a DNA test excludes a suspect who actually is the source of the genetic evidence. False negative test results may result in law enforcement excluding the actual perpetrator as a suspect in a crime, thus allowing him to remain in society and possibly commit more crimes. The laboratory error rate is essentially where human error becomes a factor in determining the validity of forensic identifications.

RMP testimony is required when DNA identification evidence is introduced (Koehler, 1993), but testimony about DNA error rates is rare (Koehler, 1996). One reason for this may be that forensic error rates have not been calculated reliably (Koehler et al., 1995; Koehler, 2013; Thompson, Kaasa, & Peterson, 2013). However, practitioners can estimate the laboratory error rate for DNA identifications through proficiency testing. In proficiency testing, the rate of erroneous DNA identifications is calculated based on how frequently forensic analysts make mistakes when they perform analyses on samples where the actual source of the DNA is known. The error rate amongst known samples is extrapolated to represent the actual error rate. (Koehler, 2013). In one instance of proficiency testing, 45 laboratories tested blood samples from known sources and, out of 223 non-matching pairs, 18 incorrect matches were reported (Koehler et al., 1995). Evaluation of a series of proficiency tests estimated that a low laboratory error rate estimate is 1-in-100 to 1-in-1000. These estimates are likely under-representative given that when labs undergo proficiency testing, they are aware that they are being tested and thus are more careful and vigorous in their sample analyses. Ideally, forensic laboratories should

not be aware of when testing is occurring and should be tested with a representative set of samples of different complexity levels to test the frequency of incorrect testing results (Koehler, 2013). Regardless of the effort analysts exert, there is no way to tell when the samples themselves are flawed; if samples are cross-contaminated or mislabeled, re-testing a reported match against the original sample will still yield the same incorrect results.

Forensic testimony should include both RMP and error rate information, but only RMP testimony is required in court (Koehler, 1993). This is a problematic oversight by the courts because, even though the rate of forensic laboratory error can only be estimated, laboratory error occurs at a much greater magnitude than errors caused by similar genetic profiles. Jurors should be able to weigh comprehensive testimony about the possibility of error in forensic test results instead of only being informed about one of the two causes of errors.

DNA Evidence in the Courtroom

Despite the possibility of errors, DNA testing is often presented as error-proof (Thompson, 2013), and, as such, is very convincing to mock jurors. Mock jurors in one study were asked to rank the reliability of five different types of evidence (Hans, Kaye, Dann, Farley, & Albertson, 2011). Sixty-four percent of participants found DNA evidence to be the most reliable type of evidence; it was considered more trustworthy than expert witnesses, police evidence, victim testimony, and eyewitness evidence (Hans et al., 2011). A second experiment found that DNA evidence was considered more reliable than suspect confessions, eyewitness identification, and victim identifications (Lieberman et al., 2008).

DNA evidence carries an illusion of infallibility, but this does not mean that jurors conceive of DNA evidence as incontrovertible evidence of guilt. For example, one mock juror experiment presented two pieces of contradictory evidence – a DNA match identifying the

defendant and an alibi witness exonerating the defendant (Golding, Stewart, Yozwiak, Djadali, & Sanchez, 2000). The introduction of an alibi witness reduced conviction rates compared to when only DNA evidence was presented, suggesting that DNA identifications are not unassailable and that contradictory evidence can instill doubt in a DNA match. In this experiment, evidence inconsistent with a DNA match reduced conviction rates, but the DNA was not presented with RMP or laboratory error rate testimony so the manipulation lacks ecological validity.

Lieberman and colleagues (2008) conducted an experiment that addressed whether jurors could understand some of the limitations of DNA evidence. They manipulated whether the laboratory doing the DNA testing had a reputation for being reliable or unreliable and jurors learned about the DNA testing through cross-examination that either focused on the DNA evidence (e.g., possibility of cross-contamination, match determination is subjective), or the expertise of the expert who testified (e.g., expertise and salary). Results indicated that when the cross-examination focused on the limitations of DNA evidence presented in a qualitative format, jurors were more likely to convict when the lab was considered reliable and less likely to convict when the lab had a reputation for being unreliable. This study focused on the caliber of the laboratory and indicated that jurors' perceptions of DNA can be influenced by cross-examination. This study did not directly extend to the possibility of laboratory error, but if jurors are sensitive to weaknesses in DNA evidence, perhaps they may also be aware of the possibility of erroneous DNA test results. It is important to directly test how jurors interpret DNA identifications and laboratory error in a narrative framework.

The Story Model

The majority of research on DNA evidence in the courtroom has focused on jurors' ability to understand and evaluate stistical information (e.g. RMP; Thompson & Schumann,

1987; Faigman & Baglioni, 1988; Goodman, 1992; Smith et al., 1996), but has neglected how DNA evidence is perceived in relation to non-forensic evidence. Forensic DNA evidence is not presented to jurors in a vacuum; it is presented in conjunction with non-forensic evidence. The Story Model of Juror Decision Making, the most widely accepted model, is an integrative framework. It states that as jurors are exposed to evidence, they incorporate each new piece of evidence into a developing narrative, or "story," representing their interpretation of the evidence (Pennington & Hastie, 1986).

Construction of the story is important since jurors are often exposed to contradictory evidence at trial and the story they choose to construct often determines which verdict the juror will ultimately endorse (Pennington & Hastie, 1988). The story model proposes that as jurors process testimony they selectively discount some of the evidence if it is inconsistent with their interpretation of the facts. For example, in one mock stabbing experiment, participants read 119 facts from multiple witnesses that contained evidence that was either ambiguous or that supported a guilty or not guilty verdict. As jurors were exposed to facts that supported different interpretations, each juror selected which facts to endorse and established a narrative interpretation of the case. After delivering a verdict, jurors rated the evidence that corresponded to their selected verdict as more important to the case than evidence that supported an alternative interpretation (Pennington & Hastie, 1988). These results suggest that jurors process testimonial inconsistencies by creating one narrative interpretation of a case and view the evidence in their selected interpretation as important to determining a verdict.

Devine and Ostrom (1985) examined how jurors evaluated evidence when one person's testimony contained information that was inconsistent with other witnesses' accounts. They presented testimony from four witnesses; three impartial witnesses provided congruent

testimony, but the testimony of the fourth witness, the defendant's sister, was inconsistent with the other testimony. The sister could be considered biased because her testimony protected her brother and indicated that he was not guilty of a crime. Evidence was either ordered by person (e.g., each witness giving all of her testimony at one time), or by evidence item (e.g., all four witnesses provided testimony about one piece of evidence until all evidence items had been reviewed). Mock jurors generally tended to discount the testimony of the biased witness, but this was particularly so when testimony was presented in witness order, likely owing to the fact that it was easier to associate the biased witness with her testimony. A follow up study manipulating the same independent variable of presentation order (i.e., evidence order versus witness order) also manipulated the perceived credibility of the one contradictory witness, so that the jurors viewed information indicating high credibility (e.g. the witness was an unbiased neighbor), low credibility (e.g. the witness was the defendant's sister), or no credibility information (Pennington & Hastie, 1992). Results indicated that jurors were more likely to discount the testimony from a low-credibility witness than a high-credibility witness. Interestingly, this finding was moderated by presentation order such that it was stronger when the evidence was presented in witness order relative to evidence order. These results indicate that jurors are less likely to discount incongruent testimony from a highly reliable source than a low reliability source. However, the existing studies only apply to eyewitness testimony and do not indicate if jurors' perceptions of a case would differ if the incongruent testimony was forensic.

Since forensic testimony is often considered highly reliable (Hans et al., 2011), it is important to evaluate how jurors respond when the only piece of forensic evidence is inconsistent with the non-forensic testimony. The novel contributions of this research are that it tests how jurors respond to DNA evidence within a narrative framework, reviews the importance

of narrative consistency between two distinct types of evidence, examines how DNA and non-forensic evidence are utilized within a trial, and looks at the relationship between forensic and non-forensic evidence in juror decision making.

Current Study

Despite the growing prevalence of forensic testimony in legal cases (Smith et al., 2011), research has not yet assessed how complex forensic evidence fits into the story model. The current project seeks to explore how jurors utilize both DNA and non-forensic evidence when selecting a verdict in two mock juror studies. Study 1 examined whether mock jurors integrate forensic evidence into the story model through participants' responses to a theft where forensic and non-forensic evidence were not always congruent. Additionally, since false positives DNA test results have more real-world consequences (e.g. wrongful convictions), Study 2 examined how the presence or absence of evidence about DNA error rates (i.e., false positive DNA test results), influenced perceptions of the evidence when non-forensic evidence was either strongly or weakly indicative of the suspect's guilt.

We hypothesized that when forensic and non-forensic evidence align, in either an inculpatory or exculpatory manner, participants would report high utilization of both the forensic and non-forensic evidence in order to decide on either conviction or acquittal, respectively. In other words, when there was a DNA match and non-forensic evidence is strong, the hypothesis was that mock jurors would report the greatest reliance on the forensic and non-forensic evidence. Alternatively, when there was no DNA match and the non-forensic evidence was weak, the hypothesis was that jurors will report the lowest utilization of both the forensic and non-forensic evidence. This project was primarily focused on the relationship between forensic and non-forensic evidence and less on the overall evaluation of the case, so, while conviction

rates and confidence in verdict accuracy are of interest, those dependent variables are not of *primary* interest.

The real focus of these experiments was on how participants evaluated evidence when forensic and non-forensic testimony were contradictory. Thus, the dependent variables of interest are to what extent mock jurors relied on the forensic and non-forensic evidence as distinct types of evidence. Two composite variables were created, DNA quality and quality of non-forensic evidence, which contain items related to either the forensic or non-forensic evidence presented at trial. Creation of these composite variables yields hypotheses about how the forensic and non-forensic testimony relate to each other. The hypothesis was that reliance on DNA quality would be driven by the strength of the non-forensic evidence. In other words, utilization of the DNA evidence would be greater when there was strong narrative evidence indicating guilt and lower when the non-forensic evidence indicated that the defendant was not guilty. The strength of the non-forensic evidence was an independent variable manipulated by the experimenters, whereas the quality of the non-forensic evidence was a composite dependent variable indicating how much participants relied on specific pieces of non-forensic evidence. The hypothesis about the quality of non-forensic evidence was that non-forensic evidence would not be influenced by whether or not there is a reported DNA match because perceptions of the narrative evidence should not be influenced by the presence or absence of DNA evidence. Logically, the presence or absence of DNA evidence should not influence how jurors interpret the narrative testimony of eyewitnesses. Investigating how DNA and narrative evidence are related can provide insight into how jurors will respond to trials with inconsistent forensic and non-forensic evidence.

Study 1

Method

Participants

Two hundred-twenty eight jury-eligible United States residents participated in this mock juror experiment. Eight participants failed attention check questions and five did not complete the experiment, so consistent with current practice, their responses are not included in the results (Oppenheimer, Meyvis, & Davidenko, 2009). The remaining sample (N = 215) was 43% male and 57% female. The age range was 18-69 years (M =35.7, SD = 12.66) with a median age of 32 years (IQR = 19). Forty-two percent of participants dichotomously identified as being religious. Regarding political ideology, on a ten-point Likert scale representing liberal (1) to conservative (10), the median was a score of four indicating that the sample tended to identify as slightly more liberal than conservative.

Participants were recruited through Amazon Mechanical Turk (AMT), which provides an online platform for "requestors" to recruit "workers" to complete brief online tasks called human information tasks (HITs). Common HITs include surveys, questionnaires, market research, and other tasks that require the workers to participate in some type of human judgment task. Using an online platform like AMT is advantageous for decision-making research because it generates a more diverse sample than studying college undergraduates (Oppenheimer et al., 2009). The range in age, SES, political orientation, and general life experience generated by online platforms tend to be more representative of the general population than undergraduate samples (Paolacci, Chandler, & Ipeirotis, 2010). The HIT for this project required that the workers meet the qualifications for jury duty in the United States, so all participants were at least 18 years of age, resided in the U.S., and did not have felony convictions. Upon completion of the task, workers

were compensated $0.75 for their participation, which is a generous AMT compensation amount (see Paolacci, Chandler, & Ipeirotis, 2010).

Procedure and Design

Participants were asked to act as jurors, read a vignette about a larceny case, and make decisions about the case. This experiment utilized a 2 (non-forensic evidence strength: strong/weak) X 2 (DNA: match reported/no match reported) between-subjects factorial design. Participants were randomly assigned to one of the four conditions.

The non-forensic evidence consisted of an eyewitness identification, financial documentation, an alibi, and motive evidence. The strong non-forensic evidence condition contained an eyewitness who was "almost certain" that he recognized the suspect, financial records indicating that the defendant had recently paid off a large debt that could not be substantiated with financial records, a co-worker who was unable to provide an alibi for the defendant's whereabouts at the time of the theft, and evidence that the defendant's 's inappropriate financial reimbursement requests would delay his growth in the company. The evidence in the weak non-forensic condition contained an eyewitness who was not confident in his identification, financial records confirming how the defendant was able to pay off a large debt, an alibi witness whose testimony made it highly improbable for the defendant to have been the perpetrator, and evidence that the defendant's inappropriate reimbursement requests were not uncommon for others in his position and his actions would not delay a promotion.

Forensic evidence was also presented in the case. The report stated that the office had been cleaned the day of the theft and investigators found the DNA of only one person on the safe. Police collected a DNA sample from the defendant. DNA testing between samples from the crime scene and the defendant indicated that the DNA either matched or did not match the

defendant. The status of the DNA evidence included information about possible errors that can occur in DNA testing.

In the "match" forensic condition, there was a reported DNA match between the DNA on the safe and the DNA provided by the defendant. Participants were told that the RMP, or the odds that such a match would occur *purely by chance*, are about 1 in 100 million. The match condition also contained information about the likelihood of false positives in DNA testing; the materials stated that,

> one type of error is a false positive, which occurs when the test declares a match between two samples when in fact the two samples do not actually match. The crime laboratory estimated that a false positive match is declared in about 1 in 100 DNA tests.

In the "no match" condition, participants are told that the DNA sample taken from the safe was not a match to the DNA taken from the suspect. The stimulus notes that the odds that the suspect's DNA and the crime scene sample *actually* do match are about 100 million to 1. The lack of a DNA match invites testimony about a different type of testing error; the stimulus included information about false negatives stating that,

> one type of error is a false negative, which occurs when the test declares that two samples do not match when in fact the two samples do actually match. The crime laboratory estimated that a false negative match is declared in about 1 in 100 DNA tests.

Participants were then asked whether they believed the defendant was guilty or not guilty of larceny. In addition to a binary verdict, participants' confidence in their decision was reported on a scale from 1 (*not at all confident*) to 9 (*extremely confident*). Another dependent variable, (What do you think the numerical likelihood is that [the defendant] committed the theft?) was

reported from 1 to 100 and was -designed to capture how likely participants thought it was that the defendant had committed the crime.

Participants answered a series of questions about how they viewed the forensic and non-forensic evidence. Questions referred to the accuracy of the eyewitness identification, the strength of the alibi evidence, the strength of the motive evidence, the accuracy of the DNA test, the likelihood that the DNA on the safe came from the defendant, and the likelihood that the DNA test returned incorrect results, among other issues. Composite scores were created for the forensic and non-forensic dependent variables. The forensic composite, DNA quality, was based on three items ($\alpha = .79$) and the non-forensic composite, quality of non-forensic evidence, was created from four items ($\alpha = .81$). Appendix A shows the items, means, standard deviations and intra-class correlations for both composite variables. The composite scores were converted to standard *z*-scores.

Results

The conviction rate was relatively low with 24.5% deciding to convict the defendant and 75.5% finding that the defendant was not guilty. As shown in Appendix B, the majority of convictions emerged from the Strong Forensic Evidence/DNA Match condition. Similarly, the likelihood of guilt estimates were low. Although the strong non-forensic evidence and DNA match condition yielded a mean estimate of guilt of 76, the estimates from the rest of the experimental conditions were all below 50.

A binary logistic regression with verdict as the dependent variable, and DNA status and evidence strength as independent variables, showed that the model was significant, ($\chi^2 = 55.34$, $df = 3$, $p < .001$), and indicated a significant main effect for DNA condition (EXP(B) = .05, Wald = 5.14, p = .02), but no significant main effect for evidence strength, (EXP(B) = .11, Wald =

3.45, *ns*), and no significant interaction, (EXP(B) = 1.66, Wald = .284, *ns*). This suggests that whether or not there was a DNA match was significantly related to how mock jurors selected their verdict; participants who heard about a DNA match were significantly more likely to convict than participants who heard that there was no DNA identification. The strength of the non-forensic evidence trended towards significance ($p = .063$), with strong non-forensic evidence tending to lead to more convictions than weak non-forensic evidence. That the only significant main effect for whether the jurors would vote to convict or not was the type of DNA test result shows the importance of researching jurors' perceptions of DNA evidence.

Participant confidence in their verdict was tested with a 2 (non-forensic evidence strength: strong/weak) X 2 (DNA: match/no match) between subjects ANOVA. There was a significant main effect for DNA status, $F(1, 213) = 4.16$, $p = .04$, $\eta^2 = .02$, but no significant main effect for evidence strength, $F(1, 213) = 1.76$, *ns*, and no significant interaction, $F(1, 213) = .18$, *ns*. Whether or not there was a reported DNA match was the only significant main effect in how confident jurors were about their verdicts. Jurors who heard that the DNA was not a match ($M = 0.14$), were more confident in their verdicts than jurors who heard there was a DNA identification, ($M = -0.14$). The strength of the non-forensic evidence did not significantly influence jurors' confidence in their verdict.

A 2 (non-forensic evidence strength: strong/weak) X 2 (DNA: match/no match) between subjects ANOVA analyzed the effect of case condition on participants' reported likelihood of guilt. There was a statistically significant main effect for evidence strength, $F(1, 213) = 43.84$, $p < .001$, $\eta^2 = .17$, and a significant effect for DNA match status, $F(1, 213) = 86.20$, $p < .001$, $\eta^2 = .29$, but there was not a significant interaction between evidence strength and DNA condition, $F(1, 215) = 2.26$, *ns*. Essentially, participants thought the defendant seemed more likely to have

committed the crime when the evidence strongly suggested he was guilty (M = 58.03) and when there was a DNA match (M = 62.53), but consistent with the hypotheses there was no interaction between evidence strength and DNA status.

DNA Quality

A 2 (non-forensic evidence strength: strong/weak) X 2 (DNA: match/no match) between subjects ANOVA was conducted to test the effect of case condition on DNA quality. The Levene's test of homogeneity of variances was not significant, so traditional F-scores will be reported. There was a statistically significant main effect for evidence strength, $F(1, 215) = 12.690$, $p < .001$, $\eta^2 = .06$, as well as a significant effect for type of DNA test result, $F(1, 215) = 8.28$, $p = .004$, $\eta^2 = .04$. There was also a significant interaction between evidence strength and DNA condition, $F(1, 215) = 8.79$, $p = .003$, $\eta^2 = .04$.

As indicated in Appendix C, when there was a DNA match, the strength of the non-forensic evidence influenced jurors' utilization of the DNA evidence. However, when there was no DNA match, the strength of the non-forensic evidence did not influence how jurors perceived the DNA evidence. In this analysis, the no DNA match condition was less important than the DNA condition because the participants were responding to how much they utilized a *lack* of DNA in their analysis of the case materials. The presence of an interaction in Appendix C indicates that, when asked about how they utilized forensic testimony in selecting a verdict, participant ratings of the DNA evidence were dependent on the nature of the non-forensic evidence. This result suggests that, consistent with the Story Model of jury decision making, jurors do not consider each piece of evidence independently. Participants' opinions of the DNA evidence were influenced by their perceptions of the non-forensic evidence.

Quality of the Non-Forensic Evidence

In evaluating the other main dependent variable, a two-factor between-subjects ANOVA tested the effects of DNA test result and evidence strength on the quality of the non-forensic evidence. Levene's statistic of homogeneity of variances was not significant, so traditional *F*-scores will be reported. There was a statistically significant main effect for evidence strength, $F(1, 215) = 47.32$, $p < .001$, $\eta^2 = .18$, and type of DNA test result, $F(1, 215) = 27.49$, $p < .001$, $\eta^2 = .12$. There was not a significant interaction between evidence strength and type of DNA test result on jurors' utilization of the non-forensic evidence, $F(1, 215) = 2.84$, *ns*.

Appendix E shows that in both conditions where the non-forensic evidence was strong and indicative of guilt, mock jurors utilized the non-forensic evidence more than participants in the conditions where the non-forensic evidence was weak, regardless of the reported DNA test results. Both when there was and was not a DNA match, non-forensic evidence strength influenced how much participants relied on the quality of the non-forensic evidence.

Discussion

The purpose of Study 1 was to analyze how mock jurors utilize DNA evidence in conjunction with narrative evidence. When the DNA and the non-forensic evidence were congruent, participants had the most extreme conviction and acquittal ratings, respectively, and the greatest confidence in those verdicts. As illustrated in Figure 1, when the non-forensic evidence was strong, a DNA identification increased the conviction rate by nearly 50% compared to when the DNA was not a match. When there was a DNA match, the conviction increased by nearly 40% when the non-forensic evidence was strong instead of weak. This suggests that jurors may weight forensic evidence more heavily than non-forensic evidence when evaluating the case facts as a whole.

Supporting the hypotheses, when the DNA evidence and non-forensic evidence were inconsistent, there was a relationship between DNA status and evidence strength. Juror reliance on DNA evidence depended on the narrative strength of the non-forensic evidence; strong non-forensic evidence led to greater reliance on a DNA match than weak non-forensic evidence.

Additionally, utilization of non-forensic evidence was dependent upon whether or not there was a DNA match. This findings shows that the jurors' thinking was illogical; the presence or absence of a DNA identification should not influence how people perceive non-forensic evidence. Since forensic and non-forensic evidence are inter-related, it is important to investigate how this relationship might be impacted by isolating the possibility that reported DNA matches may be incorrect.

Study 2

As shown in Study 1, DNA evidence is compelling to jurors and is interrelated with the narrative strength of non-forensic evidence. As such, it is important to investigate factors that influence jurors' perceptions of the probative value of DNA identification evidence and their perceptions of possible errors in DNA testing within the context of narrative evidence strength. It is essential to investigate how jurors respond to possible errors in reported DNA matches and evaluate whether the strength of the narrative evidence influences how jurors evaluate the possibility of error.

The Story Model suggests that the narrative strength and cohesion of a case will be the driving determinants of the ultimate verdict (Pennington & Hastie, 1988). As such, the story model would suggest that mock jurors who are exposed to laboratory error rate testimony would include it in their evaluation of the evidence and either incorporate it into the narrative or discount the importance of laboratory error rate testimony. Narrative information has been shown to drive decision-making more so than statistical information (Winterbottom, Bekker, Connor, & Mooney, 2008), so it is particularly important to evaluate whether or not jurors are sensitive to error rates.

The presentation of DNA evidence in Study 1 was ecologically valid; DNA matches should always include whether or not a match is reported, RMP, and the likelihood of laboratory error (i.e. reported DNA matches are always accompanied with the false positive error rate and the failure to obtain a DNA match is associated with false negative error rate testimony). Thus, in design of Study 1, it was not possible to tease apart the influence of the DNA test result from the laboratory error rate testimony. In other words, it is not clear if the effects were driven by the test result or by the possibility of laboratory error. Study 2 isolated the effect of the laboratory

error rate from the DNA test result. Since reported DNA matches and the accompanying possibility of false positive test results are objectively more important, in terms of leading to wrongful convictions, a reported DNA match with or without laboratory error rate testimony is the forensic variable of interest in this study.

To investigate how jurors utilize laboratory error rate testimony in a mock trial scenario, we manipulated the strength of the non-forensic evidence and whether a DNA match was only accompanied by the RMP or if it also included laboratory error rate evidence. By isolating the error rate testimony, we could identify group differences based on how the error rate influenced perceptions of DNA evidence. The question of interest is how jurors might change their case narratives in light of laboratory error rate testimony. For example, in a case with weak non-forensic evidence and a DNA match, will evidence about the possibility that the DNA match was a false positive test result change juror perceptions about the evidence or influence verdicts? The hypothesis is that perceptions of DNA are dependent on the consistency of the entire narrative. In other words, if the narrative of the case suggests that the defendant is guilty, DNA testimony on its own will add to jurors' construction of the model, but when the DNA match includes laboratory error rate testimony, reliance on the DNA will decrease slightly. Alternatively, when the narrative flow of the evidence is exculpatory, mock jurors will use a DNA match with laboratory error rate testimony to facilitate an acquittal whereas they may report lower reliance on all case evidence when exculpatory evidence is presented in conjunction with a DNA match, contrary to the rest of the evidence. We also hypothesized that numeracy might act as a covariate where individuals high in numeracy were more sensitive to error rate testimony than people who are low in numeracy.

Method

Participants

Jury-eligible United States residents were recruited from AMT to (N = 219) participate in a mock juror experiment. Recruited from where? If MTurk, were any precautions taken to ensure these were distinct subjects? Eleven participants failed attention check questions and four did not complete the experiment, so the remaining sample included 204 participants. The sample was 41% male and 59% female with an age range was 18-79 years (M = 36.61, SD = 13.54) and a median of 33 years (IQR = 18). On a dichotomous religiosity question, 40.6% participants identified as being religious. The political orientations of participants are as follows: 34% Independent, 33.5% Democrat, 17% Republican, 2.4% Tea Party, 1.4% Green Party and 11.8% selected none of the above. Participants were recruited through Amazon Mechanical Turk (AMT) and were paid $0.75 for their participation after completion of the task.

Procedure and Design

Participants read a vignette similar to that in Study 1. To obtain a more equally split verdict distribution than in the first study, the original stimulus was modified in two ways. First, two additional non-forensic facts were included. Secondly, the modified stimulus includes evidence about the analysis of DNA left at the scene of the crime; the DNA sample was always reported as a match and the information about the DNA match either contained or did not contain information about the laboratory error rate associated with a reported DNA match. This experiment utilized a 2 (non-forensic evidence strength: strong vs. weak) X 2 (DNA: match vs. match + error rate) between-subjects factorial design. Participants were randomly assigned to one of the four separate conditions.

The non-forensic evidence was almost identical to Study 1. A co-worker of the defendant thought she heard the defendant talking about the safe (guilty) or talking about office

furniture (not guilty) around the time of the theft and documentation about when the defendant left the office on the day of the theft that either suggested he was in (guilty) or out of (not guilty) the office were incorporated into the existing stimulus. In the strong non-forensic evidence condition, the pieces of evidence suggest that the defendant committed larceny. In the weak non-forensic evidence condition, the evidence indicates that the defendant did not commit larceny.

Similar to the forensic testimony provided in Study 1, the stimulus stated that the office had been cleaned the day of the theft and investigators found the DNA of only one person on the safe. They had also collected a DNA sample from the defendant. In the "match" forensic condition, there was a DNA match between the DNA on the safe and the DNA provided by the defendant. Participants were told that the RMP, or, the odds that such a match would occur *purely by chance* are about 1 to 100 million.

In the "match plus error rate" condition, in addition to the testimony participants received about the RMP, this condition also contained information about the likelihood of false positives in DNA testing; the materials stated that,

> one type of error is a false positive, which occurs when the test declares a match between two samples when in fact the two samples do not actually match. The crime laboratory estimated that a false positive match is declared in about 1 in 100 DNA tests.

Participants were then asked whether they believed the defendant was guilty or not guilty of larceny. In addition to a binary verdict, participant confidence in their decision was reported on a 9-point scale, with 1 representing not at all confident and 9 being extremely confident. A second dependent variable designed to capture perceptions of how likely it was that the

defendant committed the crime asked participants to provide the numerical likelihood from 1-100 that they thought the defendant committed the theft.

Participants were asked a series of questions about how they viewed the forensic and non-forensic evidence. For example, participants responded to Likert scale questions regarding the accuracy of the eyewitness identification, the strength of the alibi evidence, the strength of the motive evidence, the accuracy of the DNA test, the likelihood that the DNA on the safe came from the defendant, and the likelihood that the DNA test returned incorrect results. Composite scores were created for the forensic and non-forensic dependent variables. The four forensic items were highly inter-correlated ($\alpha = .832$) and the three non-forensic items were also inter-correlated ($\alpha = .817$). Appendix D shows the items, means, standard deviations, and intra-class correlations for the composites. The composite scores and reported means were converted to standard *z*-scores. These composite variables are the primary dependent variables because they represent the weight, as a whole, that participants put on the forensic versus non-forensic evidence determining verdicts.

Results

The conviction rate was relatively evenly split with 47.7% voting to convict the defendant and 52.3% deciding that the defendant was not guilty. As shown in Appendix F, the vast majority of convictions emerged from the Strong Forensic Evidence condition. Similarly, the likelihood of guilt estimates were split consistently based on the strength of the non-forensic evidence.

A binary logistic regression with verdict (guilty vs. not guilty) as the dependent variable indicated that the model was significant, ($\chi^2 = 126.52$, $df = 3$, $p < .001$) and detected a main effect for evidence condition, (EXP(B) = 32.34, Wald = 41.77, $p < .001$, 95% C.I. [11.18,

93.60]). Participants in the strong non-forensic evidence condition were 32.34% more likely to convict than participants in the weak non-forensic evidence conditions. There was no main effect for error rate condition, (EXP(B) = .575, Wald = .84, *ns*), and no interaction, (EXP(B) = 1.55, Wald = .239, *ns*).

A 2 (non-forensic evidence strength: strong/weak) X 2 (DNA: match/match + error rate) between subjects ANOVA was conducted to test the effect of case condition on how confident participants were in their verdict. There was no statistically significant main effect for evidence strength, $F(1, 201) = .856$, *ns*, or error rate condition, $F(1, 201) = .562$, *ns*, but there was a significant interaction between evidence strength and error rate condition, $F(1, 201) = 4.54$, $p = .034$, $\eta^2 = .02$. Gabriel's post hoc test did not reveal significant mean differences between the four conditions. However, it is possible that people in the second most confident condition, weak non-forensic evidence and DNA with the laboratory error rate, may have used the error rate testimony to justify disregarding the DNA evidence and voted to acquit; as shown in Appendix F, this group reported the lowest conviction rate. The presence of laboratory error rate testimony may have increased their confidence that the defendant was not guilty despite the incongruence of exonerating narrative evidence and a reported DNA match.

A second 2 (non-forensic evidence strength: strong/weak) X 2 (DNA: match/match + error rate) ANOVA analyzed the effect of case condition on participants' reported likelihood of guilt. There was a main effect for non-forensic evidence strength, $F(1, 201) = 208.46$, $p < .001$, $\eta^2 = .50$, but no main effect for error rate status, $F(1, 201) = .21$, *ns*), and no significant interaction, $F(1, 201) = 1.48$, *ns*). This suggests that participants' perceptions of the defendant's likelihood of guilt was primarily based on the non-forensic evidence strength and were not influenced by DNA error rate testimony. In other words, strong narrative evidence that suggests

the defendant is guilty led to higher reported likelihoods of guilt than when the narrative evidence did not strongly imply guilt. Likelihood of guilt perceptions were not influenced by error rate testimony which provided information about the possibility of an incorrect reported DNA test.

DNA Quality

A 2 (non-forensic evidence strength: strong/weak) X 2 (DNA: match/match + error rate) between subjects ANOVA was conducted to test the effect of case condition on DNA quality. The Levene's statistic of homogeneity of variances was met, so traditional *F*-scores will be reported. There was a statistically significant main effect for evidence strength, $F(1, 201) = 65.34$, $p < .001$, $\eta^2 = .24$, but not for the presence of the error rate in the DNA testimony, $F(1, 201) = 2.18$, *ns*. There was also not a significant interaction between evidence strength and DNA condition, $F(1, 201) = .012$, *ns*. In both cases where the non-forensic evidence was strongly indicative of the defendant's guilt, participants report finding the forensic evidence extremely influential, regardless of whether their experimental condition contained the possibility of a false positive error in the DNA test. There were essentially indistinguishable differences in how mock jurors responded to forensic error rate testimony when selecting their verdict.

Quality of the Non-Forensic Evidence

A two-factor between-subjects ANOVA tested the effect of case condition on quality of the non-forensic evidence. Levene's statistic of homogeneity of variances was not significant, so traditional *F*-scores will be reported. There was a statistically significant main effect for evidence strength, $F(1, 201) = 269.13$, $p < .001$, $\eta^2 = .559$, but not for the type of forensic testimony provided, $F(1, 201) = .048$, *ns*. There was not a significant interaction between evidence strength and the possibility of error rate on the reliance of the non-forensic evidence in

decision-making, $F(1, 201) = .187$, *ns*. In both conditions where the evidence was indicative of guilt, mock jurors reported relying more strongly on the non-forensic evidence than participants in the condition where the non-forensic evidence was weak. The similarity between the DNA match and DNA match with error rate conditions suggests that the forensic testimony did not have any influence on participants' responses to the non-forensic testimony.

Discussion

The aim of this study was to investigate how jurors responded to DNA error rate evidence in relation to non-forensic evidence. The only significant interaction was for how confident mock jurors were with the verdict they had chosen. It is possible that laboratory error rate testimony can influence verdict confidence when inconstant testimony is presented.

Regarding the DNA evidence, results indicated that there were essentially no differences between conditions that contained DNA or DNA plus error rate testimony. The fact that mock jurors did not use laboratory error rate testimony to reduce reliance on the non-forensic evidence when it was weak supports the idea that people are not sensitive to this type of qualifying statistical information.

Essentially, all of the results were driven by the strength of the non-forensic evidence. Unsurprisingly, when the non-forensic evidence was strong and suggested that the defendant was guilty, mock jurors were more likely to vote to convict and report a higher likelihood that the defendant committed the crime. They also relied more strongly on the DNA evidence and the non-forensic evidence than when participants were exposed to weak non-forensic evidence. These findings signify that jurors are not sensitive to the possibility of erroneous DNA test results which is problematic because, even if the non-forensic evidence is exculpatory, jurors do

not seem to be engaging critically with DNA evidence. It appears that once a match is reported, jurors rarely evaluate whether the DNA test may be flawed.

General Discussion

DNA identifications are compelling to jurors, but jurors seem to underestimate the possibility of error in reported DNA matches. Study 1 indicated that the story model of jury decision making incorporates forensic evidence. Mock jurors integrated DNA evidence into their narrative interpretations of a case. The presence or absence of a reported DNA match and the strength of the non-forensic evidence were interrelated. This suggests that mock jurors were integrating contradictory forensic and non-forensic evidence into their perceptions of a case. This research expands the existing story model work on witness testimony to include forensic evidence. It showed that forensic evidence was incorporated into jurors' narratives and became interrelated with other testimony.

Inconsistencies in the evidence led mock jurors to convict and report likelihoods of guilt that were lower than when consistent evidence suggested guilt and higher than when consistent evidence suggested innocence; this implies that mock jurors were influenced by the evidentiary inconsistencies and modified their perceptions of the evidence because of the incongruence.

However, when evaluating whether mock jurors were sensitive to laboratory error rate testimony within a case narrative, the results were troublesome. The lack of main effects for error rate testimony in Study 2 is concerning given how pervasive the influence of DNA evidence was in Study 1. DNA identification evidence is incredibly influential at trial, but the results suggested that jurors are not inherently aware of the risk of incorrectly reported DNA test results. The results suggest that error rate testimony does not influence the narrative that jurors construct and does not modify their utilization of DNA evidence. It is possible that the manipulation was not salient enough to actually influence participant perceptions of DNA, but it is also possible that jurors do not integrate error rate testimony into their perceptions of DNA. It

seems that jurors underutilize error rate testimony according to a normative model (Thompson & Schumann, 1987; Faigman & Baglioni, 1988; Goodman, 1992; Smith et al., 1996), and the results of Study 2 indicate that the error rates are overlooked in a descriptive model as well.

In Study 2, it was not clear if participants were integrating laboratory error rate testimony into their narratives; participants were not influenced by error rate testimony regardless of the strength of the non-forensic evidence, but it is not clear if they were processing, but discounting the possibility of error or if they did not understand how to think about the error rate evidence. It is possible that mock jurors understood the error rate evidence, but were not convinced that an error had occurred in the particular case they were called upon to adjudicate. It is also possible that participants did not understand the statistical RMP and laboratory error rate testimony (Taroni & Aitken, 1998). The lack of main effect for error rates testimony in Study 2 could represent mock jurors deliberately discounting the laboratory error rate or illustrate a boundary condition of the story model where complex statistical information is not incorporated into the narrative. Future research should clarify if and how jurors integrate complex statistical information into their case narratives to investigate how statistical information is incorporated into the story model. Additionally, research should tease apart whether mock jurors discounted error rate testimony or were unconvinced that it had occurred in the specific case. One way to further examine this issue it to replicate and extend these studies by investigating whether jurors are still insensitive to the possibility of erroneous DNA test results if the error rate is made more salient (e.g. error occurs in 1-in-10tests). This would reveal whether jurors are disinclined to process evidence about a DNA match analytically or whether the 1-in-100 erroneous test results was considered too unlikely for jurors to discount the DNA match.

Future research could manipulate the of evidence presentation. The Story Model would suggest that order of evidence is important and it is possible jurors might be more sensitive to weaknesses in DNA if that was the first evidence presented. This would be an interesting and informative research question, but it would not be ecologically valid. It does not make sense for a trial to begin with a DNA identification before any other evidence has been introduced to prove that a crime even occurred.

It is important to conduct more research on how, and under what conditions, jurors are sensitive to laboratory error rate testimony, especially since DNA identification evidence becomes intertwined with perceptions of non-forensic evidence. Utilization of forensic error rates should be investigated both within the broad scope of the trial, as we did in Study 2, and within the narrower context of evaluating the probative value of a specific DNA sample. At present, the frequency of false positive DNA identifications is unknown as is the rate at which those errors have facilitated wrongful convictions. Improvement in laboratory proficiency testing could generate more accurate likelihoods of error which would demonstrate whether jurors' utilization of error rates was more common when error rate testimony was based on actual proficiency testing data instead of estimates.

The usual limitations of mock trial settings apply to this article (see generally, Bornstein & McCabe, 2005). Participants responded to a brief synopsis of a legal trial with no possibility of sentencing a defendant to prison and no opportunity for jury deliberation. Therefore, these results should be interpreted tentatively until more realistic replications are conducted, but they should not be disregarded because laboratory simulations tend to generalize well to settings with more external validity (Bornstein, 1999). While the stimulus materials might lack ecological validity, previous research with extensive interviewing of mock jurors (Pennington & Hastie,

1986), suggest that jurors construct a story and evaluate evidence based on narrative consistency. This phenomenon should not change because forensic evidence has been included in the case or because the responding to the stimulus materials takes less time than being a juror on an actual.

Another limitation of this study is that the participants were AMT workers. Some criticize AMT workers as expert subjects who may not be representative of the jury pool. However, the value of AMT workers, in terms of greater variability in age, education, and general life experience, makes them a valuable sample. Additionally, since AMT workers were randomly assigned a condition, they were not in a position to make assumptions about the hypotheses of the experiments.

Since DNA and non-forensic evidence are inter-related, but jurors are not sensitive to the possibility of error, the practical implications of these experiments suggest that society is, at present, unaware that trials with weak non-forensic evidence and a DNA match may end in wrongful convictions. Additionally, these results suggest that the mere presence of a reported DNA match may inflate the perceived value given to non-forensic evidence.

DNA evidence is an incredibly valuable prosecutorial tool. It is important to consider the complicated interactions between forensic and non-forensic evidence and how the possibility of error is under-evaluated when analyzing forensic evidence. Continuing to research how jurors perceive and utilize forensic error rates could potentially reduce the overwhelming influence of reported DNA matches and reduced the likelihood of wrongful convictions.

References

Bornstein, B. H. (1999). The ecological validity of jury simulations: Is the jury still out? Law and Human Behavior, 23, 75–91. doi:10.1023/A: 1022326807441

Bornstein, B. H., & McCabe, S. G. (2005). Jurors of the absurd? The role of consequentiality in jury simulation research. Florida State University Law Review, 43, 443–467.

Devine, D. J. (2012). *Jury decision making: The state of the science*. NYU Press.

Devine, P. G., & Ostrom, T. M. (1985). Cognitive mediation of inconsistency discounting. *Journal of Personality and Social Psychology, 49*(1), 5-21.

Faigman, D. L., & Baglioni Jr, A. J. (1988). Bayes' theorem in the trial process: Instructing jurors on the value of statistical evidence. *Law and Human Behavior, 12*(1), 1-17.

Golding, J. M., Stewart, T. L., Yozwiak, J. A., Djadali, Y., & Sanchez, R. P. (2000). The impact of DNA evidence in a child sexual assault trial. *Child Maltreatment, 5*(4), 373-383.

Goodman, J. (1992). Jurors' comprehension and assessment of probabilistic evidence. *American Journal of Trial Advocacy, 16*, 361-390.

Hans, V. P., Kaye, D. H., Dann, B. M., Farley, E. J., & Albertson, S. (2011). Science in the jury box: Jurors' comprehension of mitochondrial DNA evidence. *Law and human behavior, 35*(1), 60-71.

Koehler, J. J. (1993). Error and Exaggeration in the Presentation of DNA Evidence at Trial. *Jurimetrics Journal*, *34*, 21-39.

Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review, 67*, 859-886.

Koehler, J. J. (2013). Proficiency tests to estimate error rates in the forensic sciences. *Law, Probability and Risk, 12*(1), 89-98. doi: 10.1093/lpr/mgs013

Koehler, J. J., Chia, A., & Lindsey, S. (1995). The random match probability (RMP) in DNA evidence: Irrelevant and prejudicial?. *Jurimetrics Journal, 35*, 201-219.

Lieberman, J. D., Carrell, C. A., Miethe, T. D., & Krauss, D. A. (2008). Gold versus platinum: Do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence?. *Psychology, Public Policy, and Law*, *14*(1), 27-62.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*, 37–44. doi: 10.1177/0272989X0102100105

Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science* 7(3), 203–223.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872. doi:10.1016/j.jesp.2009.03.009

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, *5*(5), 411-419.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*(2), 242-258. Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 521-533.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of personality and social psychology*, *62*(2), 189-206.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological science*, *17*(5), 407-413. doi: 10.1111/j.1467-9280.2006.01720.x

Roos, J. M. (2012). Measuring science or religion? A measurement analysis of the National Science Foundation sponsored science literacy scale 2006–2010. *Public Understanding of Science*, *0,* 1-17. doi: 10.1177/0963662512464318

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science*, *309*(5736), 892-895. doi: *10.1126/science.1111565*

Schklar, J., & Diamond, S. S. (1999). Juror reactions to DNA evidence: Errors and expectancies. *Law and Human Behavior*, *23*(2), 159-184.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, G. H. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*, 966–971.

Scurich, N. (2014). The Differential Effect of Numeracy and Anecdotes on the Perceived Fallibility of Forensic Science. *Psychiatry, Psychology and Law*, (ahead-of-print), 1-8. doi: 10.1080/13218719.2014.965293

Scurich, N., & John, R. S. (2013). Mock jurors' use of error rates in DNA database trawls. *Law and human behavior*, *37*(6), 424-431.

Skolnick, P., & Shaw, J. I. (2001). A comparison of eyewitness and physical evidence on mock-juror decision making. *Criminal Justice and Behavior*, *28*(5), 614-630. doi: 10.1177/009385480102800504

Smith, B. C., Penrod, S. D., Otto, A. L., & Park, R. C. (1996). Jurors' use of probabilistic evidence. *Law and Human Behavior*, *20*(1), 49-82.

Taroni, F., & Aitken, C. G. (1998). Probabilistic reasoning in the law: Part 1: assessment of probabilities and explanation of the value of DNA evidence. *Science & Justice*, *38*(3), 165-177.

Thompson, W. C., Kaasa, S. O., & Peterson, T. (2013). Do Jurors Give Appropriate Weight to Forensic Identification Evidence?. *Journal of Empirical Legal Studies*, *10*(2), 359-397.

Thompson, W. C., Taroni, F., & Aitken, C. G. (2003). How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic sciences*, *48*(1), 47-54.

Thompson, W. C. (2013). Forensic DNA evidence: the myth of infallibility. In S. Krimsky & J. Gruber (Eds.), *Genetic Explanations: Sense and Nonsense* (227-347). Harvard University Press.

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, *11*(3), 167-187.

Appendix A

Means, Standard Deviation, and Intra-class Correlation of Items

	M	*SD*	*ICC*			
DNA Quality						
How accurate do you think the DNA test was?	7.15	1.84	1			
How strong do you consider the DNA evidence in this case?	6.22	2.6	0.57	1		
How much of an influence did the DNA evidence in this case have on your verdict?	6.07	2.6	0.44	0.68	1	
Quality of the Non-Forensic Evidence						
How accurate do you think photocopy repairman's identification of Jason is?	3.95	2.19	1			
How likely is it that [the defendant] took money from his company to pay off his credit card?	4.88	2.49	0.57	1		
How strong do you consider the overall evidence in this case?	5.37	2.42	0.44	0.57	1	
How believable is the evidence in this case?	6.2	2.1	0.41	0.48	0.58	1

Note. Likert scales ranged from 1(low endorsement of item) to 9 (high endorsement of item).

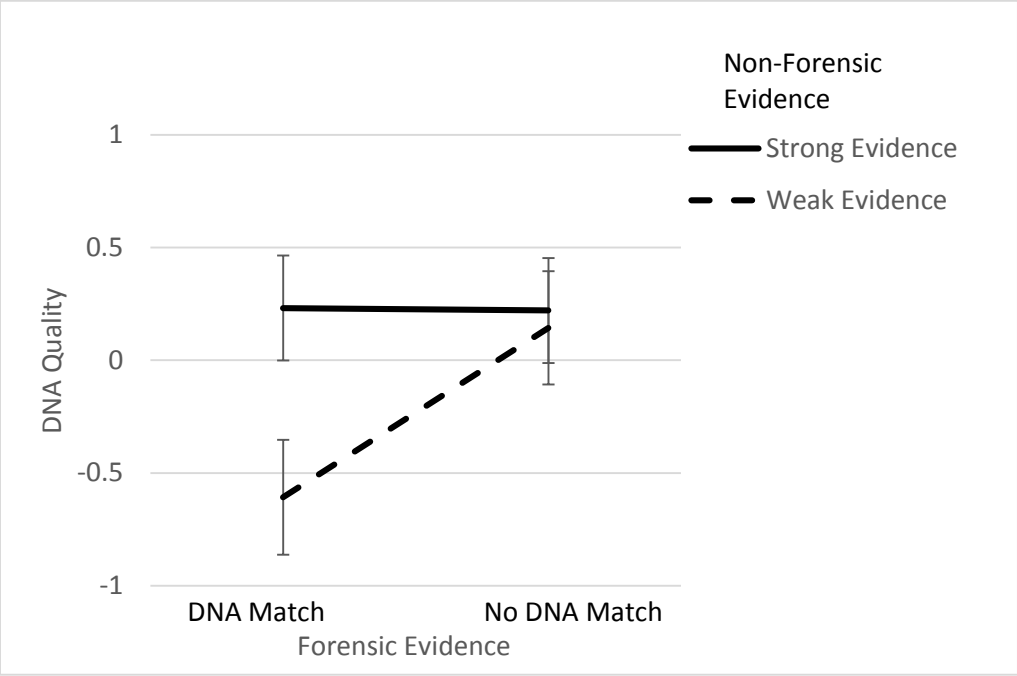
Appendix B

Table of Conviction Rates and Likelihood of Guilt by Experimental Condition

	Strong Non-Forensic Evidence			Weak Non-Forensic Evidence		
	Conviction Rate	Likelihood of Guilt	N	Conviction Rate	Likelihood of Guilt	N
DNA Match	60% (.49)	76.26 (24.11)	55	21.2% (.42)	48.81 (26.23)	53
No DNA Match	11.5% (.32)	39.81 (26.93)	52	0.04% (.19)	22.51 (21.14)	55

Note. Numbers in parentheses indicate one standard deviation. Likelihood of defendant guilt is calculated on a 0-100 scale from the question: "What do you think the numerical likelihood is that [the defendant] committed the theft?"

Appendix C



Interaction of Evidence Strength and DNA Condition Based on standardized DNA Quality Scores. Note that error bars indicate 95% C.I.

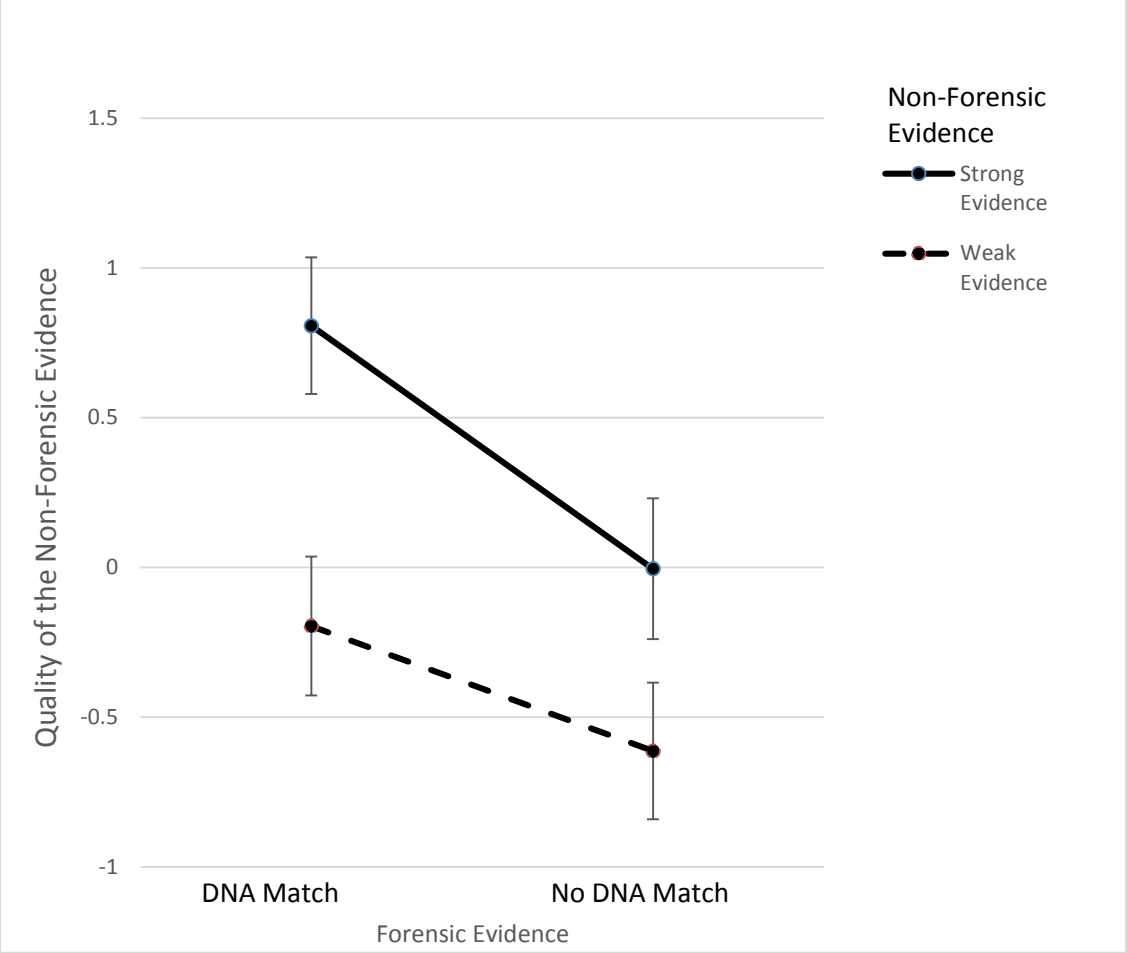
Appendix D

Means, Standard Deviations, and Intra-class Correlations for Study 2 Composites

	M	*SD*	*ICC*			
DNA Quality						
How likely is it that [the defendant's] DNA was on the safe?	7.4	1.97	1			
How accurate do you think the DNA test was?	6.84	2.36	0.74	1		
How strong do you consider the DNA evidence in this case?	5.91	2.67	0.47	0.61	1	
How much of an influence did the DNA evidence in this case have on your verdict?	5.46	2.79	0.34	0.47	0.74	1
Quality of the Non-Forensic Evidence						
How accurate do you think photocopy repairman's identification of [the defendant] is?	4.67	2.65	1			
How upset do you think [the defendant] was over the issue of his expenses getting reimbursed?	5.81	2.83	0.5	1		
How likely is it that [the defendant] took money from his company to pay off his credit card?	6.5	2.23	0.77	0.51	1	

Note. Likert scales ranged from 1(low endorsement of item) to 9 (high endorsement of item).

Appendix E



Graph of Evidence Strength and DNA Condition Based on standardized DNA Quality Scores. Error bars indicate 95% C.I.

Appendix F

Table of Conviction Rates and Likelihood of Guilt by Experimental Condition

	Strong Non-Forensic Evidence			Weak Non-Forensic Evidence		
	Conviction Rate	Likelihood of Guilt	N	Conviction Rate	Likelihood of Guilt	N
DNA Match	84.9% (.72)	84.86 (22.24)	54	14.8% (.72)	42.11 (28.63)	53
DNA Match + Error Rate	83.8% (.75)	87.39 (16.24)	54	9.1% (.58)	35.70 (27.43)	55

Notes: Numbers in parentheses indicate standard deviations. Likelihood of defendant guilt is calculated on a 0-100 scale from the question: "What do you think the numerical likelihood is that [the defendant] committed the theft?"