

# How Jurors Evaluate Fingerprint Evidence: The Relative Importance of Match Language, Method Information, and Error Acknowledgment

*Brandon Garrett and Gregory Mitchell\**

Fingerprint examiners use a variety of terms and phrases to describe a finding of a match between a defendant's fingerprints and fingerprint impressions collected from a crime scene. Despite the importance and ubiquity of fingerprint evidence in criminal cases, no prior studies examine how jurors evaluate such evidence. We present two studies examining the impact of different match phrases, method descriptions, and statements about possible examiner error on the weight given to fingerprint identification evidence by laypersons. In both studies, the particular phrase chosen to describe the finding of a match—whether simple and imprecise or detailed and claiming near certainty—had little effect on participants' judgments about the guilt of a suspect. In contrast, the examiner admitting the possibility of error reduced the weight given to the fingerprint evidence—regardless of whether the admission was made during direct or cross-examination. In addition, the examiner providing information about the method used to make fingerprint comparisons reduced the impact of admitting the possibility of error. We found few individual differences in reactions to the fingerprint evidence across a wide range of participant variables, and we found widespread agreement regarding the uniqueness of fingerprints and the reliability of fingerprint identifications. Our results suggest that information about the reliability of fingerprint identifications will have a greater impact on lay interpretations of fingerprint evidence than the specific qualitative or quantitative terms chosen to describe a fingerprint match.

## I. INTRODUCTION

Forensic evidence gathered by crime scene investigators plays an important role in many criminal cases (Durose 2005), and fingerprint impressions are the most common type of forensic evidence used to link individuals to particular crimes (Peterson et al. 2010). The popularity of fingerprint evidence is easy to understand: fingerprints are often left at crime

---

\*Address correspondence to Gregory Mitchell, School of Law, University of Virginia, 580 Massie Rd., Charlottesville, VA 22903; email: greg\_mitchell@virginia.edu. Garrett is the Roy L. and Rosamond Woodruff Morgan Professor of Law, University of Virginia; Mitchell is the Mortimer M. Caplin Professor of Law and the Class of 1948 Professor of Scholarly Research in Law, University of Virginia. Authorship order was determined alphabetically.

Simon Cole, Itiel Dror, Jennifer Laurin, John Monahan, Erin Morris, and participants at workshops at the Columbia Law School and University of Illinois College of Law provided helpful comments on a draft.

scenes, suspects can be compelled to give fingerprint impressions to compare with crime scene prints, and government agencies maintain large archives of fingerprints that can be used in crime investigations. The FBI currently maintains a computerized national databank containing over 100 million fingerprints, and every year this database is used to generate possible match candidates from the collected reference samples to aid in the investigation of thousands of crimes.<sup>1</sup>

At a criminal trial involving fingerprint evidence, a fingerprint examiner will typically testify that she followed a standard procedure for comparing prints recovered from the crime scene to reference sample prints and opine that this procedure resulted in a conclusion that the defendant was the source of the crime scene prints (see, e.g., *United States v. Baines* 2009). The ability of the fingerprint examiner to convince the jury of a fingerprint match can be crucial to the outcome of a case. In recognition of the importance of testimony by fingerprint examiners, a working group convened by the National Institute of Standards and Technology (NIST) and the National Institute of Justice (NIJ) recently promulgated standards for the production and dissemination of fingerprint examiner opinions, emphasizing that examiners should state their conclusions in a way that “will not confuse or mislead the fact-finders” (NIST/NIJ 2012).

Scholars had for some time suggested that in addition to lacking adequate scientific support, terminology commonly used by fingerprint examiners could have a significant and misleading impact on jurors (e.g., Saks 2003; Koehler & Saks 2008; Cole 2011). The National Research Council (NRC), in its recent report on forensic science, cited the “profound effect” that the use of terms like “match” and “identical” and “cannot be excluded as the source of” may have on jurors and worried that the lack of standardized language could itself lead to differences in outcomes (NRC 2009:21). The NRC called for the creation of a federal agency to establish standard scientific terminology for forensics, but no legislation has been enacted. As one concerned and candid fingerprint examiner put it after the NRC report: “We can’t say what we used to say and we can’t yet say what we will say. What do we say in the meantime?” (Cole 2011:470). While the NIST/NIJ working group was convened in part to address this concern, that group could not provide definitive guidance on the terminology that fingerprint examiners should use when testifying about a possible match because no prior research has investigated how actual or potential jurors react to and understand the testimony of fingerprint examiners.

We present the results of two studies demonstrating that different ways of presenting precisely the same fingerprint evidence can greatly affect the weight given to that evidence. We find that these differences arise primarily from the way the possibility of an identification error is addressed by the examiner, rather than from the way a match or individualization conclusion is stated. Thus, while the recent NRC report on forensic science (NRC 2009) and the NIST/NIJ report (2012) are rightly concerned that different ways of

---

<sup>1</sup>The Automated Fingerprint Identification System, or AFIS, is a national database available to federal, state, and local investigators. In addition to fingerprints, the database contains facial and tattoo photographs and descriptive information that may be used to identify persons. For AFIS information, see <[http://www.fbi.gov/about-us/cjis/fingerprints\\_biometrics/iafis/iafis](http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis)> and <[http://www.fbi.gov/about-us/cjis/fingerprints\\_biometrics/iafis/iafis\\_facts](http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis_facts)>.

presenting the same fingerprint evidence may lead to different interpretations of that evidence, their focus on avoiding the false impression of great precision in fingerprint identifications may be incomplete, or even misplaced. We find that modest and immodest individualization claims by fingerprint examiners (e.g., simple match testimony vs. testimony expressing certitude in the match and claiming the risk of error is infinitesimal) are often given the same weight by laypersons. The concern with false impressions of infallibility in fingerprint identifications, on the other hand, seems well-placed. Informing laypersons that a particular fingerprint identification could be mistaken appears to greatly reduce the weight given to fingerprint examiner testimony. Our findings suggest that courts, policymakers, and researchers should devote greater attention to developing uniform standards for accurately measuring and describing the risk of identification errors by fingerprint examiners.

## II. THE ART AND SCIENCE OF FINGERPRINT IDENTIFICATIONS

Fingerprints have been used to identify people for well over a century (Cole 2001), but the assumptions that underlie fingerprint identifications have been subjected to little scientific testing. Fingerprint examiners assume that friction ridge patterns on the skin are unique to each individual and that a partial fingerprint impression left at a crime scene can be reliably linked to an individual through microscopic comparison of friction ridge patterns and minutiae (Neumann et al. 2012). Because no objective method for assessing match probability is presently used by fingerprint examiners, such as a validated method based on statistical analyses of combinations of fingerprint features (Neumann 2012), the comparison process requires a number of subjective judgments by an examiner regarding the suitability of samples for comparison and the degree of similarity between the latent prints and reference sample prints.<sup>2</sup> This procedure can yield only qualitative individualization judgments: the prints were unsuitable for comparison, the prints matched or did not match, or the comparison produced inconclusive results. Recent studies have shown that this procedure can be distorted by motivational and cognitive factors (Dror et al. 2006; Dror & Rosenthal 2008). Although mistaken fingerprint identifications have come to light, there are few public data on the frequency with which fingerprint identifications lead to false-negative and false-positive errors (Tangen et al. 2011).

Despite the lack of an objective method for quantifying uncertainty and the dearth of data on error rates, fingerprint examiners often testify that a positive match excludes all others in the world as the source of the crime scene print and that their method is a form of rigorous science that leads to infallible or nearly infallible conclusions (Cole 2011;

---

<sup>2</sup>When we refer to the examiner's "method," we use the term colloquially rather than to assert that examiners employ a single systematic or scientific method. The NRC, in its recent report on forensic sciences, describes the analysis, comparison, evaluation, and verification (ACE-V) steps used by many fingerprint examiners as "a broadly stated framework for conducting friction ridge analyses" and states that "this framework is not specific enough to qualify as a validated method for this type of analysis" (NRC Report 2009:142). We agree with the NRC's characterization of ACE-V.

Geddes 2009, 2010; Neumann 2012).<sup>3</sup> Scholars, and even some within the fingerprint community, have long questioned the scientific basis for fingerprint examiners' claims (e.g., Cole 2001, 2004; Cummings & Midlo, 1943; Mnookin 2001, 2008; Saks 2003; Stoney & Thornton 1986; Stoney 1991), particularly the claim that the comparison process is reliable and infallible, but these criticisms have had little impact on the practices of fingerprint examiners or the judges who rule on the admissibility of their testimony (Saks 2009). The forensic science landscape changed dramatically, however, with the 2009 report of the National Research Council Committee on Identifying the Needs of the Forensic Science Community (NRC Report).

The NRC Report provides an authoritative opinion from the scientific community that a number of established forensic techniques, including individualization via fingerprint comparison, lack a rigorous scientific foundation and that there is little empirical justification for language commonly used by forensic experts to describe the likelihood that a particular suspect did or did not leave some type of forensic evidence at the scene of a crime. Professional forensic science associations and federal government agencies immediately responded to the NRC Report by changing guidelines for forensic practices and convening working groups to address the problems raised in the report (NIST/NIJ 2012). A resolution recently approved by the American Bar Association (ABA) House of Delegates encourages judges and lawyers to ensure that "experts present their testimony in a manner that accurately and fairly conveys the significance of their conclusions" (ABA House of Delegates 2012).

Unwarranted individualization claims are commonplace in known wrongful convictions (Garrett & Neufeld 2009), and the NRC Report expressed concern that the language used by fingerprint examiners and other forensic scientists may "have a profound effect on how the trier of fact . . . perceives and evaluates scientific evidence" (NRC 2009:21). Consistent with the NRC's concern, prior studies have found that jury-eligible adults have difficulty understanding expert testimony, particularly quantitative or statistical evidence (e.g., Kaasa et al. 2007; Kaye & Koehler 1991; Koehler 2001). That body of research suggested that jurors would give "unexpectedly heavy weight to the qualitative, subjective testimony typical of most forensic identification expert evidence" (McQuiston-Surrett & Saks 2009:438). However, no prior study has directly examined how jurors interpret qualitative match terminology used by fingerprint examiners.

The most closely related research examined how judges and potential jurors respond to a forensic scientist's qualitative and quantitative descriptions of the results of bite mark and hair comparisons. Conducting the first research of this kind, McQuiston-Surrett and Saks (2008) found that simple testimony from a forensic dentist that the defendant's bite pattern "matched" or was "consistent with" the bite mark found on the victim was more powerful than when the expert stated that a match was "probable" or a "reasonable scientific certainty," even though the latter terms are understood by the expert community to express stronger conclusions of association. In a follow-up study, McQuiston-Surrett and Saks (2009) found that qualitative testimony about a hair comparison in terms of a "match"

---

<sup>3</sup>NIST/NIJ calls such testimony "common practice" (2012:72).

or “similar in all microscopic characteristics” led to significantly higher estimates of the defendant as the source of the hair found at the crime scene than testimony stated in probabilistic terms, such as when an expert stated a subjective probability that a person chosen at random would have hair with the same characteristics as those found at the scene. McQuiston-Surrett and Saks (2009) further found that informing participants of limitations of the forensic expert’s method, either through cross-examination or a judicial instruction, had no effect on participants’ estimates that the defendant was the source of the hair found at the crime scene.

Koehler (2011), in a study of potential jurors’ reactions to testimony about a shoe-print match in a hypothetical criminal case, likewise found that risks of error brought out on cross-examination of a forensic expert had no effect on the persuasiveness of the expert’s testimony. Koehler (2011) did find, however, that an expert who admitted a risk of error during direct examination was less persuasive than an expert who made no such concession. Contrary to McQuiston-Surrett and Saks (2009), Koehler (2011) found no difference in the weight given to qualitative individualization testimony versus testimony in the form of a probabilistic match statistic. Aside from the important work by McQuiston-Surrett and Saks and Koehler, no further studies have been reported on nonquantified forensic testimony.

### III. STUDY 1: EXAMINING THE RANGE OF FINGERPRINT EXAMINER TERMINOLOGY

In addition to examining for the first time how laypersons interpret fingerprint examiner testimony, the current research examined a much wider range of qualitative match terminology than previously studied. Fingerprint examiners employ a range of terms and phrases to convey information about individualization via fingerprint evidence, about their methodology, and about the likelihood that their conclusion is in error (Cole 2011). Recommended terminology is found in the highly influential testimonial guidelines put forward by the International Association for Identification (IAI) and the FBI’s Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) (NIST/NIJ 2012). In criminal trials, however, examiners can vary in their adherence to such guidelines, and so we examined recent criminal trial transcripts to find examples of such variations in testimony.

Examiner testimony suggested by these guidelines and found in real trials may be placed into four broad categories: simple positive match testimony, bolstered or elaborated positive match testimony, inconclusive match testimony, and exclusion testimony. Table 1 displays examples of testimony falling within these categories. Simple positive match testimony presents an unadorned conclusion that the fingerprint evidence concluded that the defendant left the print at the scene of the crime, usually through an explicit, unqualified claim of individualization. Bolstered positive match testimony adds detail about the rigor of the fingerprint examiner’s method, about the low risk of error, or both. Some evidence suggests that forensic analysts are not often subjected to cross-examination that highlights the subjective nature of their comparisons and the risk of error (Garrett & Neufeld 2009).

Table 1: Examples of Fingerprint Examiner Terminology

**Simple Positive Match**

1. A latent fingerprint found at the scene was individualized as the left thumb of the defendant.
2. The latent fingerprint found at the scene matched the left thumb print on the ink card labeled as taken from the defendant.
3. A latent fingerprint found at the scene was made by the same individual who made the left thumb print on the ink card labeled as taken from the defendant.
4. I am confident that the defendant was the source of the latent fingerprint found at the crime scene.
5. I concluded that it is very likely that the defendant was the source of the latent fingerprint found at the crime scene.

**Bolstered Positive Match**

6. I conclude to a practical 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.
7. I conclude to a reasonable degree of scientific certainty in the field of latent fingerprint examination 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.
8. 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.
9. 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.
10. I concluded that the latent fingerprint found at the crime scene came from the same source as the left thumb print on the ink card labeled as taken from the defendant. The chance of having two individuals with the same fingerprint is one chance in 10 to the 86th power. That is more people and more fingers than are on our planet today.
11. The latent fingerprint found at the scene of the crime was individualized as the left thumb of the defendant. No two fingerprints have ever been found to be the same. The likelihood that the impression was made by a different source is so remote that it is considered a practical impossibility. Individualization is supported by the theories of biological uniqueness and permanence, probability modeling, and empirical data gained through more than one hundred years of operational experience.

**Qualified or Inconclusive Match**

12. 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.
13. 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.
14. I cannot exclude the defendant as the source of the latent fingerprint found at the scene.

**Exclusion**

15. The friction ridge impression excluded the left thumb of the defendant and originated from a different source.

NOTE: This table provides examples of testimony by fingerprint examiners. The examples were developed from a variety of sources. No. 1: "individualization" is the now-standard term for a conclusion that the print can be matched (SWGFAST 2011; NIJ/NIST 2012); No. 2: testimony reporting a "match" is the traditional conclusion reached in a latent fingerprint comparison (e.g., Trial Transcript at 403, *United States v. Baxter*, F-5928-01 (D.C. Sup. Ct. June 16, 2003); see also Cole 2007:127); No. 3: based on an FBI examiner's testimony that a print "was made by the same individual" (Trial Testimony at 105, *United States v. Lutrell Gilchrist*, No. F-2158-01 (D.C. Sup. Ct. Mar. 13, 2003); see also Cole 2007:126); Nos. 4 and 5: this type of source attribution testimony is apparently common (see Cole 2007:473); the language of No. 4 was based on testimony of an FBI analyst that "I am confident in the conclusion that I rendered" (Hearing Transcript at 124, *United States v. Faison*, 2008-CF2-16636 (D.C. Sup. Ct. May 27, 2010)); No. 8: courts have required use of "reasonable scientific certainty" language for other forensic evidence and in at least one fingerprint case and the sample language has been proposed for fingerprint cases (Cole 2011:472); No. 9: this language comes directly from current SWGFAST guidelines (SWGFAST 2011); No. 10: this language comes directly from testimony of a former FBI supervisor (Hearing Transcript at 50, *United States v. Baines*, No. 06-CR-1797 (D.N.M. Apr. 11, 2007)); No. 11: the added language regarding theoretical support for individualization comes directly from the 2009 SWGFAST guidelines. In addition, the examples were developed from information provided by counsel who have participated in criminal cases involving fingerprint evidence.

In examples of actual testimony we have obtained, when error was discussed by fingerprint examiners, it was typically to downplay the risk of error, as illustrated in the examples in Table 1, rather than to disclose risks of misidentification that the jury should take into account when weighing the evidence. Also, government experts may present inconclusive match testimony or testimony that the fingerprint evidence excluded the defendant. Since cases in which probative fingerprint evidence excludes the defendant would be less likely to go to trial, our focus here is on incriminating fingerprint evidence offered in support of the prosecution.

The NRC Report (2009) focused on the manner in which individualization conclusions are expressed, and, as Table 1 illustrates, examiners employ a range of qualitative terms that may lead to different interpretations of the same results of a fingerprint comparison. But as Table 1 also illustrates, examiner testimony often contains more information than a bare conclusion individualizing a latent print, and this additional information may impact how the conclusion is interpreted. Our first study compared the impact of these different ways of phrasing the bare conclusion and of bolstering this bare conclusion with statements about the examiner's method and its precision and reliability.

In particular, the first study examined whether different ways of phrasing a positive match caused different weights to be given to the fingerprint evidence. The standard fingerprint comparison procedure (known as Analysis, Comparison, Evaluation, and Verification, or ACE-V) results in a subjective judgment by the examiner of individualization, exclusion, or inconclusive (e.g., SWGFAST 2011). In theory, the different ways of describing an individualization conclusion should be given equal weight because the process involves no objective evaluation of the reliability or precision of the match (Neumann 2012). Research into lay meanings attached to linguistic probabilities finds consistent rank orderings of verbal quantifiers across persons but different numerical probabilities attached to particular verbal quantifiers (e.g., poor chance tends to be ranked below good chance, but the particular numerical probabilities assigned to each phrase vary widely across persons) (Dhami & Wallsten 2005; Honda & Yamagishi 2006). This research has not examined interpretations of phrases meant to convey degree of similarity, such as a "match" or "likely match" between two fingerprints. Nevertheless, we predicted that participants' beliefs that a defendant committed a crime would reflect similar weighting of fingerprint evidence regardless of the simple match terminology used when those beliefs were measured on a simple qualitative scale. However, we predicted that the numerical probabilities assigned to the likelihood that the defendant left his prints at the scene would vary more widely across individuals, making it more likely that the match terminology used would affect the weight assigned to evidence on a quantitative scale (i.e., we predicted consistency across match terms on the qualitative measure of guilt, but inconsistency across terms on the quantitative measure of the probability that the defendant left his prints at the scene).

The first study also examined whether bolstering simple match terminology with method information, claims of certitude, or low-error-risk information would increase the weight given to fingerprint evidence. On the one hand, such information should have no impact on the weight given to fingerprint evidence if people generally assume fingerprint examiners' conclusions are accurate and reliable. On the other hand, bolstering may increase the weight given to positive match testimony if people do not assume that all or

most fingerprint identifications are precise and reliable or if making the foundation for the identification explicit signals special reliability or confidence on the part of the examiner.

The first study asked three additional questions. First, does exclusion testimony or testimony about an inconclusive fingerprint comparison adversely affect the prosecution's case relative to a baseline control condition in which no fingerprint evidence was used? It is possible that the prosecution, by affirmatively addressing a lack of fingerprint evidence in its case-in-chief can blunt the impact of negative match testimony presented by a defense expert, but it is also possible that the prosecution will undercut its own case by putting on evidence of an inconclusive or negative match. We examined this possibility by including conditions in which the prosecution puts on in its case-in-chief inconclusive and negative fingerprint evidence (or an "exclusion," as an examiner would put it).

Second, do individuals differ predictably in the weight given to fingerprint evidence? Numerate individuals, or those who show greater facility working with numerical concepts, differ from innumerate individuals in their behavior on a number of judgment and decision-making tasks (Peters 2012; Peters et al. 2006), including being less trusting of verbal risk assessments (Gurmankin et al. 2004). Kaasa et al. (2007) found that mock jurors who were confident in their ability to evaluate statistical evidence used probabilistic evidence of a ballistic match more appropriately than those who were less confident in their statistical ability. We predicted that more numerate participants would appreciate the inherently probabilistic nature of fingerprint evidence regardless of the language used to convey an individualization conclusion (i.e., the ACE-V process can produce only subjective beliefs about the likelihood of a match) and would thus be less trusting of this evidence and give the evidence less weight than less numerate individuals. We also expected more numerate individuals to be less influenced by bolstering testimony that seeks to treat the fingerprinting process as nearly infallible given a greater understanding of probabilities and probabilistic processes (i.e., we predicted that bolstering would have no impact on weights assigned to the evidence by more numerate individuals but would impact innumerate individuals). In addition to numeracy, we gathered information on demographic and political differences among participants to examine whether demographic subgroups or political groups differed in their interpretations of fingerprint examiner testimony.

Third, do laypersons assume that fingerprint patterns are "unique" to individuals? In particular, laypersons may assume that not only are fingerprint patterns unique to individuals but, second, that the impressions that those fingerprints make can themselves reliably be identified as deriving from a common and single unique source. That second assumption, that accuracy in a given case can be inferred from uniqueness, has been called a "fallacy" in the scholarly literature (Koehler & Saks 2008; Cole 2004). The accuracy of an identification depends not only on the existence of identifiable differences across all persons, but also on the quality of the latent prints left at a crime scene and the reliability of the comparison method (i.e., an expert identification depends both on the reliability of the expert's method and diagnosticity of the evidence) (Kaasa et al. 2007; Schum 1994). Yet laypersons may fail to appreciate that biological uniqueness does not connote identification accuracy. If people have not considered that latent prints may not always be reliably linked to a pristine print, and instead assume that it is straightforward to assess whether a print came from a unique source, an assumption of uniqueness should increase their receptivity



to fingerprint evidence. For if one assumes that accuracy in a given case can be inferred from “uniqueness,” then the fingerprint examiner’s task is likely seen as simply determining whether this defendant left his unique marks at the scene (i.e., the reference class is a single individual). In contrast, if one does not assume that accuracy in a given case can be inferred from “uniqueness,” then the weight given to fingerprint identifications should vary with one’s assumptions about base rate similarities in fingerprint patterns (Koehler & Saks 2010).

## A. Method

### 1. Participants

Participants were recruited online through Amazon.com’s Mechanical Turk (MTurk) service ( $N = 1,252$ ; 45.9 percent female;  $M_{\text{age}} = 31.7$ ).<sup>4</sup> One group of participants received 25 cents (in U.S. dollars) for their participation ( $n = 846$ ), and a second group of participants received 50 cents (in U.S. dollars) for their participation ( $n = 406$ ). The majority of participants were from the United States ( $n = 604$ ) and India ( $n = 495$ ). We report the results for all participants because geographic location of the participant and participant payment rate had no effect on the results (i.e., the same patterns hold when we examine only the U.S. participants and when we compare the participants receiving 25 cents vs. 50 cents).<sup>5</sup>

### 2. Procedure and Materials

After their informed consent, participants provided demographic information (age, sex, education, race/ethnicity as either European American or not, and country location), described their political views on a five-point scale (from very liberal to middle of the road

---

<sup>4</sup>The use of data collected through MTurk is becoming common within psychology and the other social sciences (Bohannon 2011; Mason & Suri 2012). Buhrmester et al. (2011) found that MTurk samples were equally or more diverse than college student samples and were more representative of noncollege populations. They also found that data collected through MTurk met or exceeded the psychometric standards of published research and found, as did we, that small differences in payment rates had no effect on responses. Paolacci et al. (2010) found that MTurk participants paid at least as much attention to experiment instructions and details as participants drawn from college participant pools, and gave responses very similar to traditional participants on a series of judgment and decision-making tasks.

<sup>5</sup>Authorities in India utilize fingerprint identifications in much the same way as in the United States. In fact, the first criminal trial involving fingerprint evidence occurred in India in 1898, and in the 1890s, British Colonial Police developed and popularized the “Henry System,” the leading fingerprint classification system, which was subsequently adopted in Britain by Scotland Yard and then “spread across the world” (Cole 2001). India has had a Centralized Finger Print Bureau (CFPB) since 1955, but is in the process of integrating regional fingerprint databases and creating a national database along the lines of the FBI’s AFIS databank (Shastri 2012). For more information on the CFPB, see <<http://ncrb.nic.in/cfpb.htm>>. India is also in the process of creating a centralized fingerprint and biometrics databank (Sharma 2011), and has scanned over 200 million people as part of a process of creating national identification cards with fingerprint and iris scan information. For information on the Unique Identification Authority of India, see <<http://uidai.gov.in/>>. Twenty participants were from England, and 18 participants were from Canada; authorities in both England and Canada also rely heavily on fingerprint identifications. It is not unusual that the great majority of our participants came from the United States and India because Amazon only allows payment in U.S. dollars or Indian rupees (Mason & Suri 2012).

to very conservative), and stated which political party they typically vote for if in the United States. Participants were then randomly assigned to one of 16 conditions that presented a description of a hypothetical robbery that is the subject of a pending trial (*ns* varied from 75 to 80 participants per condition). The control condition simply described a crime for which a suspect had been arrested.

A convenience store was robbed. The robber wore a mask and used a gun. The police arrested a person who was found in the vicinity shortly after the robbery. No proceeds of the crime were found on this person, and the clerk at the convenience store has not been able to identify this person as the robber because the robber wore a mask.

No other information about the prosecution's evidence against the defendant was provided in the control condition.

The 15 other conditions provided information about fingerprint evidence collected from the scene that was compared to fingerprint samples obtained from the defendant. The 15 variations on fingerprint examiner testimony found in Table 1 were used in the 15 conditions in which fingerprint evidence information was provided to participants. For instance, in the condition providing simple individualization language (Example 1 in Table 1), participants received the following additional information.

At the trial of this case, the prosecution will present the testimony of a fingerprint examiner. The fingerprint examiner will testify as follows at trial:

I received a set of latent fingerprints taken from the crime scene. I compared these to the known fingerprints taken from the defendant on an inked card. A latent fingerprint found at the scene was individualized as the left thumb of the defendant.<sup>6</sup>

After reading the short case description, all participants were asked to rate the likelihood that the defendant committed the crime on a seven-point scale (1 = the defendant definitely is not the robber; 4 = each option is equally likely (i.e., there is a 50 percent chance the defendant is not the robber and a 50 percent chance the defendant is the robber); 7 = the defendant definitely is the robber) and were asked to rate their confidence in that judgment on a five-point scale (1 = no confidence at all; 3 = moderate confidence; 5 = complete confidence (no doubt at all)). Participants in the 15 conditions in which fingerprint evidence information was provided (i.e., all but the control condition, which contained no fingerprint information) were then asked to rate the probability that the defendant left his fingerprints at the scene of the crime on a 0–100 probability scale.

After completing the questions about the hypothetical case, participants were asked, “[d]o you believe that each person’s fingerprints are unique (i.e., do not match anyone else’s fingerprints)?”, with responses to this question either yes or no. Finally, participants completed the eight-item Subjective Numeracy Scale (Zikmund-Fisher et al.

---

<sup>6</sup>The note to Table 1 includes sources for these variations. Three of the 15 variations, for example, used the term “individualization” (one of the five “simple” variations and two of the “bolstered” variations) because it is the term adopted in professional guidelines, and because we observed its use in testimony. Other variations reflected standard language contained in professional guidelines, or variations in testimony by examiners at criminal trials.

Table 2: Dependent Measure Means by Testimony Condition (Study 1)

<i>Fingerprint Examiner Testimony Condition</i>	<i>Likelihood Defendant Was Robber (1–7 Scale) M (SD)</i>	<i>Confidence in Likelihood Judgment (1–5 Scale) M (SD)</i>	<i>Probability Defendant Left Fingerprints at Scene (0–100 Scale) M (SD)</i>
Control	3.16 (1.14)	3.42 (0.93)	—
<b>Simple positive match</b>			
1	4.28 (1.35)	3.19 (0.88)	63.67 (23.09)
2	4.54 (1.22)	3.25 (0.84)	63.84 (27.97)
3	4.34 (1.22)	3.25 (0.86)	62.48 (23.30)
4	4.40 (1.38)	3.22 (0.99)	64.99 (24.52)
5	4.36 (1.22)	3.24 (1.01)	62.40 (24.52)
<b>Bolstered positive match</b>			
6	4.52 (1.34)	3.30 (0.97)	64.92 (25.78)
7	4.40 (1.40)	2.99 (0.98)	63.71 (23.98)
8	4.26 (1.27)	3.04 (0.93)	60.28 (24.08)
9	4.36 (1.42)	3.21 (0.95)	66.03 (26.29)
10	4.57 (1.34)	3.40 (0.92)	66.53 (23.79)
11	4.56 (1.43)	3.39 (0.80)	70.04 (25.47)
<b>Qualified or inconclusive match</b>			
12	3.03 (1.13)	3.15 (1.01)	37.66 (23.17)
13	3.50 (1.18)	3.09 (1.00)	48.71 (25.57)
14	3.56 (1.02)	3.01 (0.91)	49.56 (23.46)
<b>Exclusion</b>			
15	3.06 (1.30)	3.15 (0.96)	41.96 (28.06)

NOTE: This table reports means on the dependent measures used in Study 1 by the type of testimony given by the fingerprint examiner. The control condition contained no fingerprint examiner testimony; the numbers in the Testimony Condition column correspond to the examples of testimony given by fingerprint examiners that are found in Table 1.

2007) ( $M_{\text{total numeracy score}} = 4.48$ ,  $SD = 0.87$ ,  $n = 1,147$ ;  $M_{\text{ability subscale}} = 4.52$ ,  $SD = 1.09$ ,  $n = 1,184$ ;  $M_{\text{preference subscale}} = 4.44$ ,  $SD = 0.94$ ,  $n = 1,194$ ).

## B. Results

We employed a series of planned comparisons to examine whether different ways of phrasing the fingerprint examiner's conclusion affected the weight given to the fingerprint evidence. In particular, we compared responses across the testimonial categories identified in Table 1 (e.g., average ratings on dependent measures in response to simple positive match testimony vs. bolstered positive match testimony), and we compared responses to particular forms of testimony within the testimonial categories (e.g., ratings in response to inconclusive match testimony vs. exclusion testimony).<sup>7</sup> Table 2 presents the mean ratings

<sup>7</sup>We took this approach rather than reporting all pair-wise comparisons because the latter approach would involve many comparisons, due to the number of experimental conditions, and thus inflate the risk of Type I error (falsely rejecting the null hypothesis of no difference). In contrast, an approach that controls for the family-wise error rate, such as a Bonferroni correction, would lead to underpowered tests of the null hypothesis (i.e., would make Type II errors more likely).

on each of our three main dependent measures for all 16 experimental conditions, and below we discuss the results of the planned comparisons.

Ratings of the likelihood that the defendant committed the crime and of the probability that he left his fingerprints at the crime scene differed significantly across testimonial categories,  $F(4, 1,244) = 56.16$ ,  $p < 0.001$ , partial  $\eta^2 = 0.153$ , and  $F(3, 1,169) = 49.25$ ,  $p < 0.001$ , partial  $\eta^2 = 0.112$ , respectively, but participants' confidence in their judgments of guilt did not,  $F(4, 1,242) = 2.20$ ,  $p = 0.067$ ,  $\eta^2 = 0.007$  (below we do discuss two individual differences in confidence ratings). Participants who received simple positive match testimony or bolstered positive match testimony did not differ in their judgments of the likelihood that the defendant was the robber ( $M = 4.38$  vs.  $M = 4.45$ ) or in the probabilities that they assigned to the defendant having left prints at the crime scene ( $M = 63.46$  vs.  $M = 65.23$ ),  $t(855) = -0.697$ ,  $p = 0.486$ , and  $t(857) = -1.043$ ,  $p = 0.297$ , respectively. However, participants who received qualified or inconclusive match testimony rated the likelihood that the defendant committed the robbery ( $M = 3.36$ ) and left his prints at the scene ( $M = 45.41$ ) significantly lower than those who received simple or bolstered positive match testimony, all  $t$ s  $> 8.89$  and all  $p$ s  $< 0.001$ .

Participants who received qualified or inconclusive match testimony rated the likelihood that the defendant committed the robbery significantly higher than participants who received the exclusion testimony,  $t(311) = 1.892$ ,  $p < 0.05$ , but their print probabilities did not differ significantly from those in the exclusion condition. This lack of a significant difference on the print probability dependent measure appears to have been the result of a floor effect at work in the exclusion and inconclusive match conditions, where the probability ratings did not differ significantly ( $M = 41.96$  vs.  $M = 37.66$ ,  $t(154) = -1.042$ ,  $p = 0.299$ ).<sup>8</sup> Participants reacted differently, however, to the inconclusive fingerprint examination (Row 12 in Table 2) versus the testimony by the examiner stating that the defendant could not be excluded as the source of the print (Row 14 in Table 2),  $t(153) = -3.067$ ,  $p < 0.001$  on robber likelihood and  $t(154) = -3.187$ ,  $p < 0.001$  on print probability. Thus, participants' ratings were sensitive to whether the examiner made any match at all, but when inconclusive or exclusion match testimony was given, participants' ratings on the dependent measures did not go below the baseline that seemed to be set by the simple fact that the prosecution had brought a case against this defendant for some reason.

We also compared what we viewed as the simplest positive match testimony (Example 2 from Table 1) with more elaborate versions of positive match testimony, which, a priori, we expected to have a greater influence on potential jurors by bolstering the certainty and precision of the match made by the examiner (specifically, Examples 9, 10, and 11 from Table 1). To our surprise, none of these bolstered matches produced significantly higher ratings on any of our three main dependent measures, all  $t$ s  $< 1.50$ . To verify this result, we also compared these three elaborated matches to the simple individualization testimony (Example 1 from Table 1), and again no significant differences were found on any of the dependent measures, all  $t$ s  $< 1.64$ .

<sup>8</sup>As further evidence of a floor effect, the control, inconclusive, and exclusion conditions did not differ with respect to the ratings given to the likelihood that the defendant was the robber, with all  $t$ s  $< 1.05$ .

Finally, we examined the impact of the examiner admitting that fingerprints from a person other than the defendant might also match the prints found at the crime scene (Example 13 in Table 1). The average likelihood that the defendant was the robber in this condition was significantly lower than the average rating given in the simple positive match condition,  $t(155) = 5.453$ ,  $p < 0.001$ , but did not differ from the average rating given in the control condition, where no fingerprint evidence was presented,  $t(155) = -1.813$ ,  $p = 0.072$ . Interestingly, compared to the control condition, participants' confidence in their ratings of guilt were significantly lower in the condition where a fingerprint match had been made but the prospect of another match had been made explicit,  $t(155) = 2.135$ ,  $p < 0.05$ . In other words, when the examiner expressly conceded that another person could have supplied the prints taken from the crime scene, it was as if there had been no fingerprint evidence introduced at all, and participants lost some confidence in their guilt judgments as they discounted the examiner's testimony.

To examine possible individual differences in responses, regression analyses were run for each of the dependent measures that included as explanatory variables demographic information, political information, and subjective numeracy scores, controlling for the influence of experimental condition.<sup>9</sup> With respect to ratings of the likelihood that the defendant was the robber, only self-placement on the liberalism-conservatism scale approached significance ( $\beta = 0.057$ ,  $p = 0.069$ ).

With respect to confidence ratings, race ( $\beta = 0.057$ ,  $p < 0.001$ ) and numeracy ( $\beta = 0.066$ ,  $p < 0.05$ ) accounted for significant amounts of variance. An examination of the condition means by race (coded as European American or not) revealed that those self-identifying as European American expressed less confidence in their likelihood ratings than those who did not self-identify as European American only in the condition utilizing the simple individualization terminology (Example 1 from Table 1): mean confidence of 2.95 versus 3.40,  $t(75) = -2.314$ ,  $p < 0.05$ . An examination of the condition means by median and tertile splits of subjective numeracy scores revealed no significant differences in confidence ratings across experimental conditions. Thus, the significant correlation between subjective numeracy scores and confidence ratings ( $r = 0.069$ ,  $p = 0.020$ ) indicates that those who reported higher subjective numeracy tended to express greater confidence, across conditions, in their ratings of the likelihood that the defendant was the robber.

With respect to individual differences in ratings of the probability that the defendant left his prints at the scene, race ( $\beta = -0.115$ ,  $p < 0.001$ ) and numeracy ( $\beta = 0.119$ ,  $p < 0.001$ ) again explained significant amounts of variance. Higher subjective numeracy scores were related to higher probability estimates across conditions ( $r = 0.142$ ,  $p < 0.001$ ). The difference in probability ratings between high and low subjective numeracy scorers was significantly different only in the simple match terminology condition (Example 2 in Table 1) (mean ratings of 58.00 vs. 73.00,  $t(72) = -2.32$ ,  $p < 0.05$ ). An examination of the condition means by race revealed that those self-identifying as European American gave significantly higher probability ratings than those who did not self-identify as European American in a

---

<sup>9</sup>The results of ordinary least squares regression are presented for all three dependent measures; similar results were found using ordinal regression on the robber likelihood and confidence measures.

number of conditions involving positive match or bolstered positive match testimony (Condition 1: 71.13 vs. 56.17; Condition 5: 70.76 vs. 54.00; Condition 6: 72.46 vs. 58.44; Condition 8: 66.87 vs. 55.48; Condition 9: 73.86 v. 59.15; all  $p$ s < 0.05, with Bonferroni correction applied for multiple pair-wise comparisons). In no condition did non European Americans give significantly higher probability estimates than European Americans.

Finally, the overwhelming majority of respondents believed that fingerprints were unique to individuals. Approximately 93 percent of all respondents indicated a belief in fingerprint uniqueness (1,142 out of 1,231 responses). Among U.S.-only respondents, 97 percent (581/598) indicated a belief in fingerprint uniqueness.

### *C. Discussion*

The results in Study 1 show that the particular language used by a fingerprint examiner to describe a match may not have the influence on jurors that has been widely supposed. The most exaggerated or bolstered match testimony did not lead to higher guilt ratings than the simple positive match testimony alone, but positive match testimony paired with an admission that someone other than the defendant could be the source of the match led to significantly lower guilt ratings than unqualified match testimony. These results suggest that participants were more attentive to the risk of error than to the particular language used to describe a match and suggest an asymmetry in how fingerprint matches are viewed by jurors: bolstering a match with even extravagant claims about the certainty of the match and dismissals of the likelihood that someone else supplied the prints did not increase the weight given to the match, yet simply acknowledging the possibility of a mismatch significantly undercut the probative value of the supposed match, to the point that the fingerprint evidence was fully discounted. For the mock jurors in Study 1, an unqualified match was given considerable weight regardless of the certitude with which the match was stated, but testimony by the examiner who admitted the possibility of a mistake was given no weight.

Generally, judgments of guilt as measured on an ordinal Likert scale and ratings of the probability the defendant left his prints at the crime scene as measured on a continuous probability scale were in accord: as ordinal ratings of guilt increased, probability ratings increased, and there was a surprising level of agreement on both dependent measures within testimonial categories. This agreement between the ordinal and continuous dependent measures among participants receiving similar forms of testimony suggests that there may be shared understandings of the quantitative significance of a qualitative match: when confronted with an unqualified match, participants moved slightly above the midpoint on the Likert scale, toward guilt, and assigned a probability of approximately 66 percent to the prospect of the defendant having left fingerprints at the crime scene; when confronted with an exclusion, an inconclusive result, or a qualified match, participants moved slightly below the midpoint on the Likert scale to a point indistinguishable from the baseline set in the control condition and assigned print probabilities below 50 percent. The level of agreement on both dependent measures within the positive match testimonial categories further supports the view that once an unqualified match had been made, adornment of the match information adds little probative value in the mind of potential jurors (cf. McQuiston-Surrett & Saks 2008, 2009).

Exclusion testimony by a fingerprint examiner reduced the defendant's guilt rating relative to that found in the qualified and inconclusive testimony conditions, but the exclusion did not harm the prosecution's case relative to the baseline established in the control condition (i.e., judgments of likelihood the defendant was the robber were not significantly lower than in the control condition). It may be, with respect to the exclusion, that participants assumed the prosecution had other, undisclosed evidence in order to take the case to trial, leading participants to discount this negative fingerprint evidence. That possibility raises a separate set of questions, and more work needs to be performed to understand whether jurors may assume that the prosecution has a stronger case than is apparent at trial. In addition, more work needs to be performed to understand how jurors interpret a forensic exclusion and what assumptions they may make to reduce the impact of such evidence. We do not want to make too much of a finding from a single simple hypothetical case, but this result tentatively indicates that the prosecution may be able to blunt the impact of negative fingerprint evidence by disclosing it before the defense does. The finding also suggests the potential importance, from a defense and a judicial perspective, of explaining to the jury the significance of a forensic exclusion.

Contrary to our prediction, individuals reporting greater preference for and ability with numbers gave greater weight to fingerprint evidence than subjectively less numerate individuals (where weight was measured by ratings of the probability that the defendant left his prints at the scene of the crime). Subjectively numerate individuals did not draw finer distinctions between the different forms of examiner testimony: across the board, numerate individuals seemed to be more accepting of fingerprint evidence than innumerate individuals, and those more confident in their numeracy were more confident in their guilt judgments. Had the evidence been more mathematically complex, perhaps subjective numeracy would have mattered more, or differently. And although the subjective numeracy scale has been validated (e.g., Zikmund-Fisher et al. 2007), perhaps an objective measure of numeracy (e.g., Lipkus et al. 2001) would have produced different results.

We also found differences by race, with those of European-American descent (traditionally thought of as whites) giving greater weight to the positive fingerprint evidence on the probability scale. This finding does not appear to be due to different levels of confidence in one's judgments because there was only one condition in which our two racial groups differed in confidence, with European Americans showing less confidence in their judgments in this condition. This finding may reflect greater suspicion of government evidence among minority groups, but our results do not allow us to say with confidence what mechanisms account for this result.

Finally, parties offering fingerprint evidence may benefit from a widespread assumption among jurors that no two fingerprints are alike. Although we did not conduct a representative survey of any population, almost all the participants in our large and very diverse sample reported that they believed that fingerprints are unique to individuals. This assumption, which should increase the probative value of a supposed fingerprint match, makes lay understanding of the reliability of the match process, rather than the specific language used to convey match information, central to interpretations of fingerprint examiner testimony.

#### IV. STUDY 2: A CLOSER LOOK AT THE EFFECT OF METHOD AND ERROR INFORMATION

Study 1 compared the effects of a wide range of examiner testimony and found that bolstering positive match testimony had little positive impact on probative value but qualifying a match negatively impacted the weight given to fingerprint evidence. Study 2 sought to replicate this result but also to compare systematically the effects of match certitude, method information, error information, and the manner of disclosing the possibility of examiner error.

The match certitude comparison examines whether adding qualitative language about the small likelihood of a source other than the defendant providing the fingerprint impressions would increase the weight given to simple individualization testimony. The bolstering language we tested was based on SWGFAST guidelines that permit an examiner to claim that her individualization renders another source a “practical impossibility.” In the terminology of the NIST/NIJ (2012) report, we compared a “specific individualization” claim to a “global individualization” claim.

Study 1’s results suggested that an important feature of fingerprint testimony may be how the examiner’s method, particularly its reliability, is described, but Study 1 did not test the impact of an extended description of method. Nor have prior studies provided detailed method descriptions to participants. Accordingly, for Study 2, we added to the examiner’s testimony a description of the ACE-V process based on SWGFAST’s summary of fingerprint examiner methodology. Our hypothesis was that a fairly standard method description would add weight to the fingerprint testimony by making the comparison process appear careful and precise. Since examiners do not have an explicit statistical or mathematical basis for their work and cannot point to any objective test results to support their conclusions, the description of the method is likely to focus on the uniqueness of fingerprints and the standard procedure to be followed carefully by the examiner, implicitly downplaying the subjective judgments involved in this method and highlighting the confidence of the examiner in the conclusions reached.

Of course, because the ACE-V process is “based largely on human interpretation” (NRC Report 2009:139), it is susceptible to human error. The NRC Report (2009) noted a series of subjective decisions in the fingerprint comparison process, including judgments about the difference between a “discrepancy” versus a “distortion” and about the number of differences that are disregarded as “explainable.” Crucially, the NRC Report noted that it is an “unrealistic” and “scientifically implausible” assumption that a method could have a zero error rate (NRC 2009:142–43). Shortly after that report was released, the IAI issued guidance that examiners should no longer claim a zero error rate (IAI 2010), and the NIST/NIJ report (NIST/NIJ 2012) likewise recommended that examiners not claim a zero error rate and that they describe steps taken to minimize sources of error.

While fingerprint examiners are now cautioned by the leading professional associations against claiming infallibility, it is not clear what they should say about the prospect of error because there is little research on error rates in general and few examiners have undergone proficiency testing (NIST/NIJ 2012). To examine how different admissions of



error by the fingerprint examiner may affect the weight given to the fingerprint evidence, we compared the impact of two types of error statements by an examiner: one statement, based on examples of actual examiner testimony, acknowledged the abstract possibility of error but then claimed near infallibility on the part of this examiner;<sup>10</sup> another statement admitted that examiners can make errors and that it therefore must be conceded that the identification in this case could be in error.<sup>11</sup> In light of the mixed findings on error disclosure between Koehler (2011) and McQuiston-Surrett and Saks (2009), we further tested whether error statements would have an effect when brought out on direct examination versus cross-examination.

### A. Method

#### 1. Participants and Design

Participants were recruited online through Amazon.com's MTurk service and were paid 50 cents for their participation ( $N = 689$ ; 57.8 percent female;  $M_{\text{age}} = 33.2$ ). Participation in this study was restricted to U.S. residents. Participants were randomly assigned to a control condition or to one of 16 conditions produced by a  $2 \times 2 \times 2 \times 2$  full factorial between-subjects design: two levels of match certitude (simple individualization testimony vs. bolstered individualization testimony), two levels of method information (information about fingerprint examiner method provided or not), two levels of error information (risk of error discounted vs. risk of error acknowledged), and two means of error disclosure (error information provided during direct examination or during cross-examination).

#### 2. Procedure and Materials

After giving informed consent, participants provided demographic information (age, sex, education, race/ethnicity as either European American or not, and country location), described their political views on a five-point scale (from very liberal to middle of the road to very conservative), and stated which political party they typically support. All participants then received a short description of a hypothetical crime and its suspect.

*a. Control Condition.* Participants assigned to the control condition received the following description.

A convenience store was robbed. The robber wore a mask and used a gun. The police arrested a person who was found in the vicinity shortly after the robbery. No proceeds of the crime were found on this person, and the clerk at the convenience store has not been able to identify this

<sup>10</sup>Simon Cole describes cases in which the fingerprint examiner described the risk of error as purely "theoretical" or "infinitesimal" (Cole 2011:472).

<sup>11</sup>Given the dearth of data on error rates associated with the ACE-V process, we did not commit our hypothetical examiner to any particular error rate; rather, he simply admits that an error in the individualization conclusion is possible.

person as the robber because the robber wore a mask. The robber did not fire the gun and dropped the gun when running out of the store after the robbery.<sup>12</sup>

Participants assigned to one of the experimental conditions received the same description, plus information about the prosecution's fingerprint evidence, which varied depending on the level of the match certitude, method information, error information, and error disclosure variables.

i. Match certitude. Those participants in the simple individualization condition received the following additional information.

Fingerprint impressions were found on the handle of the gun. At the trial of this case, the prosecution presented the testimony of a fingerprint examiner who examined those prints. The fingerprint examiner testified as follows at trial:

I collected a set of fingerprint impressions from the gun found at the crime scene and determined that one fingerprint was suitable for comparison. I compared this fingerprint to the known fingerprints taken from the defendant on an inked card. The fingerprint found on the gun was individualized as the right thumb of the defendant.

Participants in the bolstered individualization condition received the above information and were also told that “[t]he likelihood the impressions were made by a different source is so remote that it is considered to be a practical impossibility.” This bolstering language was taken from current SWGFAST guidelines for fingerprint examiner testimony.

ii. Method information. Participants in the method information condition received the following description of fingerprint comparisons at the beginning of the fingerprint examiner's direct testimony.

To understand my testimony, it may help to have some basic information about fingerprints and the fingerprint identification process.

Fingerprints are permanent. They are formed before birth, and they remain the same until after death, barring deep scarring. The underside of our fingers and hands and feet are covered with raised skin, called friction skin, which is usually covered with a thin film of perspiration or oil. When the finger or hand touches an item, a reproduction of those ridges is left by means of that perspiration or oil. That reproduction is called a latent print. Studies have found that no two people have the same ridge arrangement anywhere on their fingers or hands.

When a fingerprint examiner receives evidence from investigating officers, the examiner develops the print with powders or chemicals to make it visible, and then photographs the print. The photograph is then studied using a standard process known as ACE-V, which stands for Analysis, Comparison, Evaluation, and Verification.

In the ACE-V process, the examiner first determines whether the print is suitable for comparison, based on whether there is adequate information in it. If so, then the examiner makes a side-by-side comparison between the print found at the scene and a known print taken from a suspect.

---

<sup>12</sup>Based on feedback from participants in Study 1, who noted that our failure to specify the location of the prints left open the possibility that the defendant had been a recent customer in the convenience store, we slightly altered the hypothetical to make clear that the recovered prints came from the gun used in the robbery.

The examiner evaluates the degree of similarity between the crime scene print and the known print and can conclude that the prints are the same, different, or that the comparison is inconclusive. Finally, if the examiner decides that the prints came from the same person, the prints are passed to another examiner, to verify that conclusion.

This abbreviated description was based on the more detailed SWGFAST description of the ACE-V process. Participants in the no-method-information condition did not receive the above information before receiving the examiner's conclusion regarding a positive match.

iii. Error information and error disclosure. Participants in the discounted-error condition were told that the examiner stated the following: "While there is always the possibility of human error in making an identification based on fingerprints, I have never made a mistake. The chances are infinitesimal that the pattern in the print found at the scene of the crime could have come from someone other than the defendant."<sup>13</sup> Participants in the error-acknowledged condition were told that the examiner stated the following: "Recent studies have found that fingerprint examiners do sometimes make mistakes about the source of a fingerprint found at a crime scene. It is possible that the defendant was not the source of the print found at the scene of this robbery."

iv. Error disclosure. Participants were told that the fingerprint examiner made one of the above error statements as either part of his direct examination or during cross-examination.

*b. Dependent Measures.* After reading the case information, participants were asked to complete the same dependent measures used in Study 1 to assess perceived weight of the evidence against the defendant (i.e., likelihood the defendant committed the crime on a seven-point scale, confidence in likelihood judgment on a five-point scale, and, in all conditions but the control condition, the probability defendant left fingerprints at the crime scene on a 0–100 probability scale). These measures appeared on the same webpage as the case description.

Finally, on a separate webpage, participants were asked whether they believed fingerprints were unique (same question as used in Study 1) and were asked to rate the reliability of criminal identifications based on fingerprint comparisons (six-point scale: 1 = not reliable at all; midpoints were not labeled; 6 = extremely reliable). Participants were also asked to respond to two questions designed to measure aversions to Type I and Type II errors: "How serious an error is it for the criminal justice system to convict an innocent person?" and "How serious an error is it for the criminal justice system to fail to convict a guilty person?" Responses to these questions were given on a six-point scale (1 = not a serious error; midpoint was not labeled; 6 = an extremely serious error). The last question in the experiment asked participants to report the frequency with which they watched any of the

---

<sup>13</sup>This claim that fingerprint comparison error rates are "infinitesimal" comes from a real case (Cole 2011:472).

crime scene investigation (CSI) television shows (six-point scale: never, rarely, once a month, a few times a month, once a week, a few times a week).

## B. Results

Analysis of variance in the mean ratings of the likelihood that the defendant was the robber found a main effect only for the error information variable,  $F(1,631) = 27.49$ ,  $p < 0.001$ , partial  $\eta^2 = 0.042$ . When the examiner downplayed the risk of error, the defendant's likelihood of guilt was rated significantly higher than when the examiner admitted that examiners can make errors and an error in the identification was possible ( $M_{\text{error discounted}} = 5.41$  vs.  $M_{\text{error acknowledged}} = 4.91$ ). There were no other significant main or interaction effects on the qualitative measure of evidentiary weight.<sup>14</sup> Compared to the strength of the case without fingerprint evidence (i.e., the control condition), even the weakest fingerprint evidence (i.e., the condition in which the risk of error was acknowledged only on cross) led to significantly higher judgments that the defendant likely committed the crime (all  $t$  tests comparing the control condition to each experimental condition were significant at  $p < 0.05$ , even after applying the Bonferroni correction for multiple comparisons) (see Table 3 for mean scores on the dependent measures for all 17 conditions).

Providing participants with bolstered match testimony increased confidence in their judgments about the defendant's likely guilt ( $M_{\text{simple individualization}} = 3.50$  vs.  $M_{\text{bolstered individualization}} = 3.63$ ,  $F(1, 629) = 4.17$ ,  $p < 0.05$ , partial  $\eta^2 = 0.007$ ), but when the examiner acknowledged the risk of error, as opposed to downplaying that risk, participant confidence dropped ( $M_{\text{error discounted}} = 3.69$  vs.  $M_{\text{error acknowledged}} = 3.43$ ,  $F(1, 629) = 16.05$ ,  $p < 0.001$ , partial  $\eta^2 = 0.025$ ). No other main or interaction effects were found on the confidence measure.

When the examiner acknowledged the possibility of error in his identification, mean ratings of the probability that the defendant left his fingerprints at the crime scene decreased ( $M_{\text{error discounted}} = 76.10$  vs.  $M_{\text{error acknowledged}} = 68.09$ ,  $F(1, 633) = 18.54$ ,  $p > 0.001$ , partial  $\eta^2 = 0.028$ ). No other significant effects were observed on the print probability measure.<sup>15</sup>

As in Study 1, very few individual differences were observed. Regression analyses on each of the dependent measures that included the range of individual difference measures obtained found only that older participants tended to give higher ratings of the likelihood that the defendant committed the robbery ( $\beta = 0.11$ ,  $p < 0.05$ ), those who watched CSI

<sup>14</sup>The timing of the error disclosure approached significance,  $F(1, 631) = 3.32$ ,  $p = 0.07$ , with the error acknowledgment that was brought out during cross-examination leading to the lowest likelihood ratings (i.e., failing to admit the risk of error until cross-examination reduced the weight given to the examiner's testimony more than volunteering this risk on direct). The interaction of error information with method information also approached significance,  $F(1, 631) = 3.11$ ,  $p = 0.08$ , with the addition of method information counteracting the effects of acknowledging a risk of error.

<sup>15</sup>The main effect for the method information variable approached significance ( $F(1,633) = 3.46$ ,  $p = 0.06$ ), with higher probability ratings associated with the examiner giving a method introduction before expressing his match conclusion.

Table 3: Dependent Measure Means by Testimony Condition (Study 2)

<i>Fingerprint Evidence Condition</i>	<i>Likelihood Defendant Was Robber (1–7 Scale) M (SD)</i>	<i>Confidence in Likelihood Judgment (1–5 Scale) M (SD)</i>	<i>Probability Defendant Left Fingerprints at Scene (0–100 Scale) M (SD)</i>
Control	3.30 (.94)	3.15 (0.95)	—
<b>Simple individualization</b>			
No method information			
EDD	5.41 (1.24)	3.56 (0.87)	75.41 (22.38)
EDC	5.38 (1.10)	3.52 (1.04)	74.03 (22.16)
EAD	4.93 (1.03)	3.33 (0.73)	68.34 (21.57)
EAC	4.66 (1.28)	3.34 (0.79)	63.02 (20.63)
Method information			
EDD	5.44 (1.23)	3.71 (0.81)	77.27 (20.96)
EDC	5.37 (1.12)	3.65 (0.80)	71.12 (27.39)
EAD	5.20 (1.13)	3.41 (0.89)	74.71 (22.26)
EAC	5.11 (1.23)	3.42 (0.76)	69.34 (26.69)
<b>Bolstered individualization</b>			
No method information			
EDD	5.51 (1.03)	3.66 (0.67)	74.27 (23.87)
EDC	5.48 (1.30)	3.85 (0.76)	78.57 (20.06)
EAD	4.93 (1.37)	3.48 (0.78)	65.95 (27.28)
EAC	4.59 (1.12)	3.39 (0.92)	63.34 (23.31)
Method information			
EDD	5.20 (1.35)	3.70 (0.83)	76.56 (26.16)
EDC	5.51 (1.36)	3.85 (0.82)	81.56 (19.59)
EAD	5.21 (1.17)	3.63 (0.71)	74.38 (22.69)
EAC	4.68 (1.35)	3.46 (0.87)	65.61 (29.35)

NOTE: This table reports means on the dependent measures used in Study 2 by the type of testimony given by the fingerprint examiner. The control condition contained no fingerprint examiner testimony. Participants received either simple individualization testimony or bolstered individualization testimony, and they received either no method information or information about the method used by the fingerprint examiner. In addition, participants read that the fingerprint examiner either discounted or acknowledged the risk of error during direct testimony or cross-examination (EDD = error risk discounted on direct; EDC = error risk discounted on cross; EAD = error risk acknowledged on direct; EAC = error risk acknowledged on cross).

shows more often had greater confidence in their guilt judgments ( $\beta = 0.08, p < 0.05$ ), and those who were more concerned about false convictions gave lower estimates of the probability that the defendant left his prints at the crime scene ( $\beta = -0.08, p < 0.05$ ).<sup>16</sup>

Also as in Study 1, participants in Study 2 overwhelmingly believed that fingerprints are unique to individuals (94.5 percent, or 651 out of 689, responded affirmatively). Participants also rated the reliability of fingerprint identifications as high ( $M = 4.43$  on a six-point scale). There were no differences in ratings of the reliability between participants who were given the method information and those who were not (i.e., adding method

<sup>16</sup>The results of ordinary least squares regression are presented for all three dependent measures; similar results were found using ordinal regression on the robber likelihood and confidence measures.

information did not increase perceptions of the reliability of the method), suggesting a high baseline belief in reliability rather than an effect of the experimental manipulation.<sup>17</sup>

### C. Discussion

The most consistent effect in Study 2 involved the error information variable: when the examiner admitted that fingerprint examiners sometimes make mistakes and that the identification in this case could thus be wrong, participants reduced their judgments about the likelihood the defendant committed the crime, reduced their estimates of the probability that the defendant left his prints at the crime scene, and had less confidence in their guilt judgments. These effects were found regardless of the certitude with which a positive match conclusion was stated, whether method information was provided, and whether the possibility of error came out on direct or cross (although the trend in the means was for revelation of the risk of error during cross-examination to have a more depressing effect on judgments of guilt). In short, when an examiner admitted the undeniable fact that fingerprint examiners can make identification errors, and that the identification at hand was thus not foolproof, significantly less weight was given to the fingerprint evidence. This result replicates Study 1's finding that qualifying a positive match by admitting the prospect of error undercuts the probative value of fingerprint evidence.

The strength of the match language had no independent effect on participant beliefs about the likelihood of the defendant being the robber or having left his prints at the scene, also replicating Study 1's finding of no impact from bolstering simple match testimony. However, participants were significantly more confident in their guilt judgments when given the bolstered match testimony. That greater confidence may make their opinions more resistant to change during jury deliberations.

Adding method information had no independent effect on beliefs about the likelihood the defendant was the robber or the probability the defendant left fingerprints at the scene. The effect of method information on print probability estimates approached significance, with the addition of method information increasing estimates of the probability that the defendant left prints at the scene, but this result should be viewed cautiously given its tentative nature. Participants exposed to the examiner who testified on direct that his method was reliable and then acknowledged on cross a possible misidentification rated the general reliability of fingerprint identifications the lowest. Thus, our results suggest that an examiner who claims infallibility on direct will be viewed skeptically after a cross that elicits error-risk concessions, but an examiner who on direct describes her method in reasonable

<sup>17</sup>General reliability ratings were marginally sensitive to the combination of error information and method information ( $F(1, 633) = 3.77, p = 0.053$ ), and the main effect for timing of the error disclosure (direct vs. cross) approached significance ( $F(1, 633) = 3.32, p = 0.07$ ). An examination of mean reliability ratings by condition revealed that participants who received error information rated the reliability of fingerprint identifications to be lower, but participants who received method information during direct examination and then an admission of a risk of error during cross-examination tended to give the lowest reliability ratings: the contradiction between the confident claims on direct and the admission on cross of less-than-perfect reliability of the examiner's method may have affected general views about the reliability of fingerprint identification. Because of these possible experimental effects on the general reliability measure, scores on the general reliability measure were not used as a moderator or individual difference variable.

terms, including acknowledging some risk of error, may be able to limit the negative impact of an effective cross-examination or contrary fingerprint evidence presented by the defense.

There were few individual differences in Study 2, and we did not replicate our finding from Study 1 that persons of different races viewed the fingerprint evidence differently. We did find, interestingly, that more frequent viewing of forensic science television dramas increased participant confidence in their evaluations of the fingerprint evidence, suggesting that the empirical debate over the impact of media on evidential expectations should continue (see, e.g., Tyler 2006; Shelton 2008). Finally, confirming the results from Study 1, the overwhelming majority of respondents in Study 2 believed that individuals have unique fingerprints.

## V. GENERAL DISCUSSION

In two large, diverse samples of adults, we found that a fingerprint examiner's testimony that a suspect's fingerprints matched latent prints recovered from a crime scene significantly increased participants' beliefs in the guilt of this suspect, but the particular terminology used to convey this positive match information had little impact on participants' judgments about the guilt of the suspect. Both studies found no effect for different ways of framing an individualization conclusion or bolstering that conclusion, such as with method information or claims that it was a practical impossibility that another person could be the source of the crime scene prints.

These results indicate that while the legal and forensics community may be rightly concerned about the manner in which forensic conclusions are expressed, modest testimonial conclusions may be just as problematic as overstated conclusions. The results also call into question the effectiveness of proposals that aim to reduce the weight given to fingerprint testimony by having examiners replace exaggerated claims of certainty and specificity with unadorned claims that simply state that a positive match has been made. Our results suggest that participants encoded examiner testimony as a dichotomous variable—as “match” or “no match”—with elaboration by the examiner on the strength of the match or certainty of the conclusion having little impact once the testimony was coded as a match.<sup>18</sup>

Our basic finding that different ways of framing an examiner's match conclusion did not matter should be viewed in conjunction with our finding that people generally view

---

<sup>18</sup>Such an encoding process would be consistent with fuzzy-trace theory (e.g., Hans & Reyna 2011; Reyna & Brainerd 1995), which assigns primary importance in judgments and decision making to the “gist” meaning extracted from observed information and leads to categorical judgments of the kind we observed across the experimental conditions with respect to the impact of positive, inconclusive, and negative match conclusions. However, fuzzy-trace theory would suggest that when the experimental task called for finer-grained judgments, as with our quantitative measure of evidentiary weight, then verbatim representations of the examiner's testimony should have played a greater role in judgments. Yet we found few differences on the quantitative measure of evidentiary weight. Future studies should more closely examine how examiner testimony is encoded and the relative impact of gist and verbatim representations of examiner testimony on quantitatively measured evidentiary weights and on categorical jury verdicts.

fingerprint identifications as reliable and believe that fingerprints are unique sources of identification. Thus, even the simplest positive match testimony can have a great impact on jurors: testimony that fingerprints “matched” appears to be taken as a strong statement of individualization that needs no bolstering. The bolstering language appears to have just reaffirmed what laypeople already believed about fingerprint evidence.

Our studies also shed light on a possibly more effective way of qualifying fingerprint examiner conclusions: when the fingerprint examiner admitted that his method is not foolproof and that his conclusion in this case could be in error, that disclosure had a significant negative impact on the evidence. This approach appears to have been more effective than qualifying the match conclusion because it contradicted preexisting beliefs about the reliability of fingerprint identifications and put fingerprint identifications in a new light. Future studies should examine more closely lay beliefs about the fingerprint identification process so that courts and parties can develop methods for separating fact from fiction in jurors’ consideration of fingerprint evidence.

Our finding that bringing out the risk of error on cross had an equal or greater impact on lay interpretations of the examiner’s testimony than voluntary disclosure during direct examination contradicts the findings of Koehler (2011) and McQuiston-Surrett and Saks (2009). These mixed results indicate that the impact of cross-examination may depend greatly on how the information that is revealed during cross-examination relates to lay assumptions about a particular type of evidence and how that information compares to claims made by the witness during direct examination. We did not explore the impact of proficiency information that was tailored to the particular examiner testifying, but our results suggest that such error-rate information and how it is conveyed could greatly affect the weight given to the examiner’s testimony. Our results also suggest that cross-examination of an overconfident examiner that garners an admission of fallibility may negatively impact jurors and, conversely, that forensic analysts may be wise to acknowledge their fallibility on direct to avoid such negative effects.

In addition, the impact of exclusion evidence and inconclusive match testimony should receive more scrutiny. Perhaps our participants appropriately ignored testimony about an inconclusive match by rating the strength of evidence in this case the same as the control case, but arguably our participants gave the exclusion testimony too little weight by not rating the strength of evidence in that case lower than the control case. It is possible that laypersons generally discount exclusion evidence, but the weight given to such an exclusion will likely depend on the whether a credible explanation exists for how this defendant could be both guilty and not the source of fingerprints. There are certainly troubling examples of serious criminal cases in which jurors apparently discounted forensic findings excluding criminal defendants who later had their innocence confirmed by DNA testing (Garrett & Neufeld 2009), but only future studies can determine how common such discounting is and whether our result was simply an artifact of our method, which did not control for participants’ assumptions about the other evidence that the government might have had against the defendant.

As with any research involving mock cases, our studies have limitations. We did not provide full case details, we focused exclusively on fingerprint evidence, and we did not utilize group deliberations, jury instructions, or attorney arguments. The participants were



not asked to reach a verdict. Our goal was to examine the marginal effects of fingerprint evidence, compared to the control case, on laypersons generally, and to examine the marginal effects of different ways of packaging a fingerprint examiner's conclusions. These marginal effects may be easily overwhelmed in actual trials by other evidence or by the demeanor and personal characteristics of the examiner who testifies. We consciously chose to examine a crime in which no one was physically harmed to minimize the threat of outcome bias. A crime involving harm to a victim may elicit strong motives to convict, which may in turn color the way evidence is interpreted and assimilated (e.g., Mazzocco et al. 2004).

Notwithstanding these limitations, our studies shed new light on issues raised in the NRC Report (2009) and the NIST/NIJ Report (2012) and raise new issues for study. Our results suggest that the precise terminology used by fingerprint examiners to convey an individualization conclusion may have relatively little impact on the weight given to that identification, even when that terminology makes exorbitant claims about the certainty of the identification. Thus, the focus by the NRC and the NIST/NIJ taskforce on specific match terminology may be misplaced.

Instead, a focus on error-rate information may be warranted. Our results support Jennifer Mnookin's recommendation that "[w]hat courts really ought to consider requiring, in many cases, as a minimum prerequisite to admissibility is simply much better error rate information about examiners' abilities in practice" (Mnookin 2010:1243). Our research demonstrates that admissions by fingerprint examiners that their method is not infallible and that a person other than the defendant could have supplied the crime scene prints can significantly affect the weight given to a fingerprint examiner's testimony. Given the indisputable fact that fingerprint examinations are not infallible and that someone other than the defendant may always be the source of crime scene prints, there is no justification for claims of infallibility or near infallibility by fingerprint examiners. Such claims are not seen as expected braggadocio that is easily ignored by jurors; rather, such claims confirm many people's preexisting beliefs about fingerprint evidence and portray examiner testimony in an overly positive light. As one fingerprint examiner put it recently, "when we say, 'I am 100% certain of my conclusion,' we might mean that we have conducted a careful examination, reached the best conclusion possible with the data available, and that we would not have reported that conclusion unless we were confident that we had done our work well. But what does the jury hear? They hear, 'I'm an expert, and I'm telling you that this conclusion is fact and cannot possibly be wrong'" (Eldridge 2012:8).

Finally, greater attention should be paid to the problem of how to appropriately express opinions across a host of forensics for which no probabilistic basis for rendering conclusions exists. The organizations developing such standards should focus on more than just the expert's bare conclusion and whether that conclusion is stated in qualitative or quantitative terms. Exactly how a forensic expert communicates an individualization conclusion may be much less important than the information the expert provides about the reliability of the process used to reach that conclusion, the risk of error, and the degree to which that information confirms or contradicts assumptions that jurors already hold about the reliability of fingerprint evidence.

## REFERENCES

- American Bar Association House of Delegates (2012) *Revised Resolution 101C*. Available at <[http://www.abanow.org/wordpress/wp-content/files\\_flutter/13285463222012mm101c.pdf](http://www.abanow.org/wordpress/wp-content/files_flutter/13285463222012mm101c.pdf)>.
- Bohannon, J. (2011) "Social Science for Pennies," 334 *Science* 307.
- Buhrmester, M., T. Kwang, & S. D. Gosling (2011) "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?" 6 *Perspectives on Psychological Science* 3.
- Cole, S. A. (2001) *Suspect Identities: A History of Fingerprinting and Criminal Identification*. Cambridge, MA: Harvard Univ. Press.
- (2004) "Grandfathering Evidence: Fingerprint Admissibility Rulings from *Jennings* to *Llera Plaza* and Back Again," 41 *American Criminal Law Rev.* 1189.
- (2007) "Where the Rubber Meets the Road: Thinking About Expert Evidence as Expert Testimony," 52 *Villanova Law Rev.* 803.
- (2011) "Splitting Hairs? Evaluating 'Split Testimony' as an Approach to the Problem of Forensic Expert Evidence," 33 *Sidney Law Rev.* 459.
- Cummings, H., & C. Midlo (1943) *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics*. Philadelphia, PA: Blakiston Co.
- Dhami, M. K., & T. S. Wallsten (2005) "Interpersonal Comparison of Subjective Probabilities: Toward Translating Linguistic Probabilities," 33 *Memory & Cognition* 1057.
- Dror, I., D. Charlton, & A. Peron (2006) "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications," 156 *Forensic Science International* 74.
- Dror, I., & R. Rosenthal (2008) "Meta-Analytically Quantifying the Reliability and Biasability of Forensic Experts," 53 *J. of Forensic Sciences* 900.
- Durose, M. R. (2005) *Census of Publicly Funded Forensic Crime Laboratories*. Available at <<http://bjs.ojp.usdoj.gov/content/pub/pdf/cpffcl05.pdf>>.
- Eldridge, H. (2012) "'I Am 100% Certain of My Conclusion.' (But Should the Jury be Certain?)," March–April *Evidence Technology Magazine* 8.
- Garrett, B. L., & P. J. Neufeld (2009) "Invalid Forensic Science Testimony and Wrongful Convictions," 95 *Virginia Law Rev.* 1.
- Geddes, L. (2009) "Forensic Evidence Goes on Trial," February 28 *New Scientist* 6.
- (2010) "Fingerprints to Go Scientific at Last," March 20 *New Scientist* 10.
- Gurmankin, A. D., J. Baron, & K. Armstrong (2004) "The Effect of Numerical Statements of Risk on Trust and Comfort with Hypothetical Physician Risk Communication," 24 *Medical Decision Making* 265.
- Hans, V. J., & V. F. Reyna (2011) "To Dollars from Sense: Qualitative to Quantitative Translation in Jury Damage Awards," 8 *J. of Empirical Legal Studies* 120.
- Honda, H., & K. Yamagishi (2006) "Directional Verbal Probabilities: Inconsistencies Between Preferential Judgments and Numerical Meanings," 53 *Experimental Psychology* 161.
- International Association for Identification (July 16, 2010) IAI Resolution 2010-18. Available at <[http://www.swgfast.org/Resources/100716\\_IAI\\_Resolution\\_2010-18.pdf](http://www.swgfast.org/Resources/100716_IAI_Resolution_2010-18.pdf)>.
- Kaasa, S. O., T. Peterson, E. K. Morris, & W. C. Thompson (2007) "Statistical Inference and Forensic Evidence: Evaluating a Bullet Lead Match," 31 *Law & Human Behavior* 433.
- Kaye, D. H., & J. J. Koehler (1991) "Can Jurors Understand Probabilistic Evidence?" 154(1) *J. of the Royal Statistical Society, Series A* 75.
- Koehler, J. J. (2001) "When Are People Persuaded by DNA Match Statistics?" 25 *Law & Human Behavior* 493.
- (2011) "If the Shoe Fits They Might Acquit: The Value of Forensic Science Testimony," 8 *J. of Empirical Legal Studies* 21.
- Koehler, J. J., & M. J. Saks (2008) "The Individualization Fallacy in Forensic Science Evidence," 61 *Vanderbilt Law Rev.* 199.
- (2010) "Individualization Claims in Forensic Science: Still Unwarranted," 75 *Brooklyn Law Rev.* 1187.

- Lipkus, I. M., G. Samsa, & B. K. Rimer (2001) "General Performance on a Numeracy Scale Among Highly Educated Samples," 21 *Medical Decision Making* 37.
- Mason, W., & S. Suri (2012) "Conducting Behavioral Research on Amazon's Mechanical Turk," 44 *Behavioral Research* 1.
- Mazzocco, P. J., M. D. Aliche, & T. L. Davis (2004) "On the Robustness of Outcome Bias: No Constraint by Prior Culpability," 26 *Basic & Applied Social Psychology* 131.
- McQuiston-Surrett, D., & M. J. Saks (2008) "Communicating Opinion Evidence in Forensic Identification Sciences: Accuracy and Impact," 59 *Hastings Law J.* 1159.
- (2009) "The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear," 33 *Law & Human Behavior* 436.
- Mnookin, J. L. (2001) "Fingerprint Evidence in an Age of DNA Profiling," 67 *Brooklyn Law Rev.* 13.
- (2008) "The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate," 7 *Law, Probability & Risk* 127.
- (2010) "The Courts, the NAS, and the Future of Forensic Science," 75 *Brooklyn Law Rev.* 1209.
- National Institute of Justice and National Institute of Standards and Technology (NIJ/NIST) (2012) *Latent Print Examination and Human Factors: Improving the Practice Through a Systems Approach*. Available at <[http://www.nist.gov/customcf/get\\_pdf.cfm?pub\\_id=910745](http://www.nist.gov/customcf/get_pdf.cfm?pub_id=910745)>.
- National Institute of Standards and Technology & National Institute of Justice (February 17, 2012) *Latent Print Examination and Human Factors: Improving the Practice Through a Systems Approach*. Available at <[http://www.nist.gov/manuscript-publication-search.cfm?pub\\_id=910745](http://www.nist.gov/manuscript-publication-search.cfm?pub_id=910745)>.
- National Research Council Committee on Identifying the Needs of the Forensic Science Community (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press.
- Neumann, C. (2012) "Fingerprints at the Crime-Scene: Statistically Certain, or Probable?" 9 *Significance* 21.
- Neumann, C., I. W. Evett, J. E. Skerrett, & I. Mateos-Garcia (2012) "Quantitative Assessment of Evidential Weight for a Fingerprint Comparison. Part II: A Generalisation to Take Account of the General Pattern," 214 *Forensic Science International* 195.
- Paolacci, G., J. Chandler, & P. G. Ipeirotis (2010) "Running Experiments on Amazon Mechanical Turk," 5 *Judgment & Decision Making* 411.
- Peters, E. (2012) "Beyond Comprehension: The Role of Numeracy in Judgments and Decisions," 21 *Current Directions in Psychological Science* 31.
- Peters, E., D. Västfjäll, P. Slovic, C. K. Mertz, K. Mazzocco, & S. Dickert (2006) "Numeracy and Decision Making," 17 *Psychological Science* 408.
- Peterson, J., I. Sommers, D. Baskin, & D. Johnson (2010) *Final Report. The Role and Impact of Forensic Evidence in the Criminal Justice Process*. Available at <<https://www.ncjrs.gov/pdffiles1/nij/grants/231977.pdf>>.
- Reyna, V. F., & C. J. Brainerd (1995) "Fuzzy-Trace Theory: An Interim Synthesis," 7 *Learning & Individual Differences* 1.
- Saks, M. J. (2003) "The Legal and Scientific Evaluation of Forensic Evidence (Especially Fingerprint Expert Testimony)," 33 *Seton Hall Law Rev.* 1167.
- (2009) "The Past and Future of Forensic Science and the Courts," 93 *Judicature* 94.
- Schum, D. A. (1994) *The Evidential Foundations of Probabilistic Reasoning*. New York: Wiley.
- Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) (2011) *Standards for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint)*. Available at <[http://www.swgfast.org/documents/examinations-conclusions/111026\\_Examinations-Conclusions\\_1.0.pdf](http://www.swgfast.org/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf)>.
- Sharma, A. (2011) "Govt Planning Database on 1.5 Crore Criminals," October 28 *India Today*. Available at <<http://indiatoday.intoday.in/story/home-ministry-plans-national-database-on-criminals/1/157679.html>>.
- Shastri, P. (2012) "Experts Advocate Nationalized Fingerprint Database," March 27 *Times of India*.
- Shelton, D. E. (2008) "The 'CSI Effect': Does it Exist?" 259 *NJJ J.* 1.

- Stoney, D. A. (1991) "What Made Us Ever Think We Could Individualize Using Statistics?" 31 *J. of the Forensic Science Society* 197.
- Stoney, D. A., & J. I. Thornton (1986) "A Critical Analysis of Quantitative Fingerprint Individuality Models," 31 *J. of Forensic Sciences* 1187.
- Tangen, J. M., M. Thompson, & D. J. McCarthy (2011) "Identifying Fingerprint Expertise," 22 *Psychological Science* 995.
- Tyler, T. R. (2006) "Viewing CSI and the Threshold of Guilt: Managing Truth and Justice in Reality and Fiction," 115 *Yale Law J.* 1050.
- United States v. Baines*, 573 F.3d 979 (10th Cir. 2009).
- United States v. Faison*, No. 2008-CF2-16636 (D.C. Super. Ct. Feb. 19, 2010).
- Zikmund-Fisher, B. J., D. M. Smith, P. A. Ubel, & A. Fagerlin (2007) "Validation of the Subjective Numeracy Scale (SNS): Effects of Low Numeracy on Comprehension of Risk Communications and Utility Elicitations," 27 *Medical Decision Making* 663.