Lay Interpretation of Fingerprint Examiner Testimony: The Relative Importance of Match Language, Method Information and Error Acknowledgement

Brandon Garrett and Gregory Mitchell

University of Virginia

Author Note

Brandon Garrett, School of Law, University of Virginia; Gregory Mitchell, School of Law, University of Virginia. Authorship order was determined alphabetically.

John Monahan provided helpful comments on a draft.

Correspondence concerning this article should be addressed to Gregory Mitchell, School of Law, University of Virginia, 580 Massie Road, Charlottesville, VA 22903.

E-mail: greg_mitchell@virginia.edu
Abstract

Fingerprint examiners use a variety of terms and phrases to convey information about a supposed match between a defendant’s fingerprints and fingerprint impressions collected from a crime scene. Two studies examined 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 conclusion—whether simple and imprecise or detailed and claiming near certainty—had little effect on participants’ judgments about the guilt of a suspect. Admitting the possibility of error decreased the weight given to the evidence regardless whether the admission was made during direct examination or cross-examination, but providing information about the method that fingerprint examiners use to make their comparisons reduced the impact of this admission of possible identification error. We found few individual differences in reactions to the evidence across a wide range of participant variables, and we found widespread agreement regarding the uniqueness of fingerprints and the reliability of fingerprint identifications. The results suggest that the background information provided 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.

Keywords: forensic science, fingerprints, individualization, expert testimony, jury decision-making

Word count: 12,169
Lay Interpretation of Fingerprint Examiner Testimony: The Relative Importance of Match Language, Method Information and Error Acknowledgement

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). This popularity is easy to understand: fingerprint evidence is often left at crime scenes, suspects can be compelled to give fingerprint impressions for comparison 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 databank containing over 100 million fingerprints, and this database is used to generate possible match candidates from the collected reference samples to aid in the investigation of thousands of crimes every year.¹

When fingerprint-based identifications are used in a criminal trial, 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 that this procedure resulted in an individualization of the defendant as the source of the crime scene prints (see, e.g., U.S. v. Baines, 2009). Where fingerprint evidence is an important part of the government’s case linking a defendant to a crime, the ability of the fingerprint examiner to inform and influence the jury will be crucial to the outcome of the case. A federal agency working group convened to examine the future use of latent fingerprint evidence noted disagreement as to how conclusions should be expressed and

¹ The Integrated Automated Fingerprint Identification System, or IAFIS, 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 IAFIS information, see http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis and http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis_facts.
called for “ongoing attention” to the problem of how to express conclusions in a way that “will not confuse or mislead the fact-finders” (NIST/NIJ, 2012). Although the particular way that a fingerprint examiner describes her method and its results may affect the interpretation and weight given to fingerprint evidence, 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.

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 microscropic 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. This procedure can only yield 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 is potentially subject to distortion by motivational and cognitive factors (Dror, Charlton & Peron, 2006; Dror & Rosenthal, 2008). Although mistaken fingerprint identifications have come to light, there is little public data on the frequency with which
fingerprint identifications lead to false-negative and false-positive errors (Tangen, Thompson & McCarthy, 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 (Geddes, 2009, 2010; Neumann, 2012, Cole 2011).² Scholars have long questioned the scientific basis for fingerprint examiner’s claims (Cole, 2004; Mnookin, 2008), particularly the claim that the comparison process is reliable and infallible, but these academic 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

---

² NIST/NIJ (2012) calls such testimony “common practice” (p. 72).
“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” (p. 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, Peterson, Morris & Thompson, 2007; Kaye & Koehler, 1991; Koehler, 2001). 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. McQuiston-Surrett and Saks (2009) found that qualitative testimony about the hair comparison in terms of a “match” or “similar in all microscopic characteristics” led to significantly higher estimates that the defendant was the source of the hair found at the crime scene than testimony stated in probabilistic terms, such as when an expert stated the 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
LAY INTERPRETATION OF FINGERPRINT TESTIMONY

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 shoeprint match in a hypothetical criminal case, likewise found that risks of error brought out on cross-examination of the 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 and testimony in the form of a probabilistic match statistic. Aside from the important work by McQuiston-Surrett, Saks and Koehler, no further studies have been reported on non-quantified forensic testimony.

**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, their methodology, and the likelihood that their conclusion is in error (Garrett, 2011). Much of this 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).

Examiner testimony may be placed into four broad categories: simple positive match testimony, bolstered 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 individualized the defendant to 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). In examples of 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 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. Our focus here is on fingerprint evidence offered in support of prosecution given the predominance of this use of fingerprint evidence.

[Insert Table 1 about here]

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 simple positive match caused different weights to be given to the fingerprint evidence. The standard fingerprint comparison method (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); different ways of describing an individualization conclusion should thus be given equal weight because the method 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 across persons (e.g., poor chance tends to be ranked below good chance, but the particular numerical probabilities assigned to each phrase vary widely) (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 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, Baron & Armstrong, 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 method can only produce subjective beliefs about the likelihood of a match) and would thus be less trusting of this evidence than innumerate individuals and would give the evidence less weight than innumerate 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 lay persons assume that fingerprint patterns are “unique” to individuals? An assumption of uniqueness should increase lay receptivity to fingerprint evidence. For if one assumes uniqueness, then the fingerprint examiner’s task is likely to be conceived 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 uniqueness, then the weight given to fingerprint identifications should vary with one’s assumptions about base rate similarities in fingerprint patterns (Koehler & Saks, 2010). Furthermore, lay people may fail to appreciate that biological uniqueness does not connote identification accuracy, for the accuracy of an identification depends not only on the existence of identifiable differences across all persons but also on the quality of 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).

Method

Participants

Participants were recruited through Amazon.com’s Mechanical Turk (“MTurk”) service ($N = 1252; 45.9\%$ female; $M_{age} = 31.7$). One group of participants received 25 cents (in U.S. dollars) for their participation ($n = 846, 39\%$ female; $M_{age} = 30.5$). A second group of participants, which was restricted to U.S. residents, received 50 cents (in U.S. dollars) for their participation ($n = 406; 60\%$ female; $M_{age} = 34.3$). The great majority of participants from outside the U.S. resided in India ($n = 495$). Geographic location and payment rate had no effect on any of the results.

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 to very conservative), and stated which political party they typically vote for if in the U.S. 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 ($n$’s 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 (No. 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."

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 (that is, there is a 50% chance the defendant is not the robber and a 50% 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 to indicate whether or not they believed that each person’s fingerprints were unique (responses to this question were either yes or no). Finally, participants completed the eight-item Subjective Numeracy Scale (Fagerlin et al., 2007) ($M_{\text{total numeracy score}} = 4.48, SD = .87, N = 1147$; $M_{\text{ability subscale}} = 4.52, SD = 1.09, n = 1184$; $M_{\text{preference subscale}} = 4.44, SD = .94, n = 1194$).

Results

Participants differed in their ratings of the likelihood that the defendant committed the robbery and in their estimates of the probability that the defendant left fingerprints at the crime scene depending on the way in which the fingerprint testimony was presented (Table 2 presents mean results for the three dependent measures by condition). But across experimental conditions, participants did not differ in the confidence with which they made those likelihood ratings (though we do note below two individual differences in confidence ratings).

[Insert Table 2 about here]

When rating the likelihood that the defendant was the robber using the seven-point ordinal scale, participants who received simple and bolstered positive match testimony rated that likelihood as significantly greater than those participants in the control, qualified/inconclusive match, and exclusion conditions. There were no differences in likelihood ratings within any of these testimony categories, however. Most notably, different ways of phrasing a positive match did not lead to different likelihood ratings, and bolstering the simple positive match testimony did not lead to higher likelihood ratings than the simple positive match testimony alone. Also, the exclusion testimony did not lead to lower likelihood ratings than those found in the control condition, suggesting a floor effect based on an assumption by participants that the government must have other, undisclosed evidence to support its prosecution of the defendant.
When rating the probability that the defendant left his prints at the crime scene using the probability scale, there was greater variation across conditions, as predicted from research on inter-individual differences in the functions used to transform verbal quantities into numerical probabilities. We observed three levels of probability ratings: the exclusion and inconclusive match conditions resulted in significantly lower probability ratings than the simple and bolstered positive match conditions; the two qualified match conditions (numbers 12 and 13 in Table 1) and three of the simple positive match conditions resulted in intermediate probability ratings; the simple individualization condition (number 1 in Table 1) and the bolstered positive match conditions resulted in the highest probability ratings.

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. With respect to ratings of the likelihood that the defendant was the robber, only self-placement on the liberalism-conservatism scale approached significance ($\beta = .057, p = .069$).

With respect to confidence ratings, race ($\beta = .057, p = .000$) and numeracy ($\beta = .066, p = .034$) 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 number 1 from Table 1): mean confidence of 2.95 versus 3.40, $F(1,75) = 5.36, p = .023$. 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 ($\beta = 0.119, p = 0.000$) again explained significant amounts of variance. Higher subjective numeracy scores were related to higher probability estimates across conditions ($r = 0.142, p = 0.000$). The difference in probability ratings between high and low subjective numeracy scorers was significantly different only in the simple match terminology condition (number 2 in Table 1) (mean ratings of 58.00 vs. 73.00, $F(1, 72) = 5.38, p = 0.023$). 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 number of conditions involving positive match or bolstered positive match testimony (condition number 1: 71.13 vs. 56.17; condition number 5: 70.76 vs. 54.00; condition number 6: 72.46 vs. 58.44; condition number 8: 66.87 vs. 55.48; condition number 9: 73.86 vs. 59.15; all $p$’s < .05, with Bonferroni correction applied for multiple pairwise 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% of all respondents indicated a belief in fingerprint uniqueness (1142 out of 1231 responses). Among U.S. only respondents, 97% indicated a belief in fingerprint uniqueness (581 out of 598 responses).

**Discussion**
The particular type of positive match terminology used by fingerprint examiners did not affect the weight given to fingerprint evidence when that weight was measured on a qualitative scale. This result held true even when simple positive match testimony was bolstered by claims about the rigor and reliability of the examiner’s method and conclusion. In all of these positive match conditions, the fingerprint evidence significantly increased judgments of the likelihood that the defendant committed the robbery above the baseline established in the control condition, where no detail about the prosecution’s case was provided.

Likewise, there were no differences in the weight given on the qualitative scale to qualified match testimony, inconclusive match testimony, and even exclusion testimony. None of these conditions—not even the condition in which the fingerprint evidence was said to exclude the defendant—harmed 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). Perhaps participants assumed the prosecution must have had some other, undisclosed evidence in order to take the case to trial and were thus willing to discount this negative fingerprint evidence, particularly where the prosecution brought this negative evidence forward. More work needs to be performed to understand how jurors interpret a forensic exclusion and what assumptions they may make that reduce the impact of such evidence. We do not want to make too much of a finding from a single simple hypothetical case, but our result tentatively indicates that the prosecution may be able to blunt the impact of negative fingerprint evidence by disclosing such evidence before the defense does so. 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.
As predicted, the weight given to the fingerprint examiner’s testimony varied more when the weight of evidence was measured on the probability scale. Participants exposed to some forms of bolstered positive match testimony gave the highest estimates of the probability that the defendant left his fingerprints at the crime scene, simple positive match testimony produced the next highest probability estimates, and the exclusion, qualified and inconclusive testimony produced lower probability estimates. These results indicate that the participants were attending to, and were sensitive to, the particular way the fingerprint examiner framed conclusions.

However, as noted above, these different probability estimates did not result in qualitative differences in judgments that the defendant was the robber. When combined with other evidence at a trial, these different probability weights might lead to different outcomes on the usual qualitative judgment offered at trial—guilty or not guilty—but in our simple scenario, the probability differences between simple positive match testimony and bolstered positive match testimony did not matter qualitatively.

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 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. 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, and European-Americans showed 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.

It is noteworthy that participant location exerted no effect on any of our results. This absence of geographic differences may be due to the fact that our non-U.S. participants resided predominantly in India, which has a criminal justice system in which fingerprint evidence is commonly used. In fact, the first criminal trial involving fingerprint evidence occurred in India in 1898 (Cole, 2001), and India is in the process of creating a centralized fingerprint and biometrics databank much like the FBI’s IAFIS databank (Sharma, 2011).

Finally, parties offering fingerprint evidence may benefit from a widespread assumption among jurors that no two fingerprints are alike, and that they are “unique” to each individual. Although we did not conduct a representative survey of any population, almost all of the participants in our large and very diverse sample reported that they believed that fingerprints are unique to individuals. This assumption may partially account for the moderately high confidence found in participants across all conditions for participants’ judgments that the defendant was the robber: if fingerprints are individually unique, then reliable match testimony should be strong evidence of the defendant’s connection to the crime. Thus, lay understanding of the reliability of the match process, rather than the specific language used to convey match information, becomes crucial to interpretations of fingerprint examiner testimony.
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 some forms of bolstering or qualifying match testimony can affect 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 results suggested that an important feature of live fingerprint testimony is a description of the method used by examiners, but study 1 did not test the impact of an extended description of method as part of the testimony. 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 method 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 method is “based largely on human interpretation” (NRC Report, 2009, p. 139), it is susceptible to human error. The NRC Report (2009) noted a series of subjective decisions in the fingerprint comparison process, including the difference between a “discrepancy” versus a “distortion” and 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 Report, 2009, pp. 142-143). Shortly after that report was released, the IAI issued guidance that examiners should no longer claim a zero error rate (IAI, 2010), while the report from the working group of the National Institute of Standards and Technology/National Institute of Justice (“NIST/NIJ”) indicated examiners should not claim a zero error rate, should be familiar with the literature on error rates, and should describe steps taken to minimize sources of error (NIST/NIJ, 2012).

While fingerprint examiners are now cautioned against claiming infallibility, it is not clear what they should say about the prospect of error and error rates because there is not adequate research on error rates in general nor for many fingerprint examiners in particular (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 the examiner; another statement admitted that examiners can make errors and that it

---

3 Simon Cole describes cases in which the fingerprint examiner described the risk of error as purely “theoretical” or “infinitesimal” (Cole, 2011, p. 472).
therefore must be conceded that the identification in this case could be in error.\textsuperscript{4} 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.

\textbf{Method}

\textbf{Participants and Design}

Participants were recruited through Amazon.com’s MTurk service and were paid 50 cents for their participation ($N = 689$; 57.8 \% 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 $2 \times 2 \times 2 \times 2$ between-subjects full factorial design: two levels of match certitude (simple individualization testimony versus 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 versus risk of error acknowledged) and two means of error disclosure (error information provided during direct examination or during cross-examination).

\textbf{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 apprehension of a suspect.

\textsuperscript{4} Given the dearth of data on error rates associated with the ACE-V method, 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.
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 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.5

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

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.

---

5 Based on feedback from participants in study 1 noting that our failure to specify the location of the prints left open the possibility the defendant had been a customer, we slightly altered the hypothetical to make clear the recovered prints came from the gun used in the robbery.
**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. 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 method. Participants in the no method information condition did not receive the above information before receiving the examiner’s conclusion regarding a positive match.

**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.” 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.”

**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.

**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, probability defendant left fingerprints at the crime scene on 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 (1-6 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

---

6This qualification that fingerprint comparison error rates are “infinitesimal” has been recently offered in litigation (Cole 2011, p. 472).
frequency with which they watched any of the 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).

**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 \( F(1,631) = 27.49, p = .00 \). When the examiner downplayed the risk of error, 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 \( (M_{\text{error discounted}} = 5.41 \text{ vs. } M_{\text{error acknowledged}} = 4.91) \). The timing of the error disclosure approached significance \( (F(1, 631) = 3.32, p = .07) \), with the error acknowledgement 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 \( (F(1, 631) = 3.11, p = .08) \), with the addition of method information counteracting the effects of acknowledging a risk of error. There were no other significant main or interaction effects on the qualitative measure of evidentiary weight. 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 \( p < .05 \) 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 \( (F(1, 629) = 4.17, p = .04; M_{\text{simple individualization}} = 3.50 \) vs. \( M_{\text{bolstered individualization}} = 3.63 \), but when the examiner acknowledged the risk of error, as opposed to downplaying that risk, participant confidence dropped \( (F(1, 629) = 16.05, p = .00; M_{\text{error discounted}} = 3.69 \) vs. \( M_{\text{error acknowledged}} = 3.43 \)). 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 \( (F(1, 633) = 18.54, p = .00; M_{\text{error discounted}} = 76.10 \) vs. \( M_{\text{error acknowledged}} = 68.09 \)). The main effect for the method information variable approached significance \( (F(1,633) = 3.46, p = .06) \), with higher probability ratings associated with the examiner giving a method introduction before expressing his match conclusion. No other significant effects were observed on the probability measure.

As in study 1, very few individual differences were observed. Regression analyses on each of the dependent measures including 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 \( (\beta = .11, p = .01) \), those who watched CSI shows more often had greater confidence in their guilt judgments \( (\beta = .08, p = .04) \), 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 \( (\beta = -.08, p = .03) \).

Also as in study 1, participants in study 2 overwhelmingly believed that fingerprints are unique to individuals \( (94.5\%, \text{ or } 651 \text{ out of } 689, \text{ responded affirmatively}) \). Participants also
rated the reliability of fingerprint identifications as high (M = 4.43). 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 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. However, general reliability ratings were sensitive to the combination of error information and method information ($F(1, 633) = 3.77, p = .053$), and the main effect for timing of the error disclosure (direct vs. cross) was marginally significant ($F(1, 633) = 3.32, p = .07$). 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.

**Discussion**

In study 2, we found that strength of the match language had no independent effect on participant beliefs about the likelihood the defendant was the robber or on participants’ estimates of the probability that the defendant left prints at the scene. 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.

---

7 Because of these possible experimental effects on the general reliability measure, scores on this measure were not used as a moderator or individual difference variable.
We also found that 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. However, 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. 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 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.

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 and their estimates of the probability he left his prints at the scene, and participants 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.

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 (see Tyler, 2006). Finally, confirming the results from study 1, the overwhelming majority of respondents in study 2 believed that individuals have unique fingerprints.

**General Discussion**

In two large, diverse samples of adults, we found that a fingerprint examiner’s testimony that a suspect’s prints 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 of our studies found no effect on a qualitative measure of evidentiary weight for different ways of framing a positive match conclusion or bolstering that conclusion in various ways, 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. We did find that in the first study that bolstering a positive match conclusion increased evidentiary weight assigned to fingerprint evidence on a quantitative scale, but in our second study bolstering with method information had only a marginal effect on even the quantitative measure of evidentiary weight and bolstered match certitude had no effect on this quantitative measure.

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 bring into question proposals that aim to reduce the weight given to fingerprint testimony by having examiners replace exaggerated claims of certainty and specificity with more cautious claims that still indicate 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. 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. When the task called for finer-grained judgments, as with our quantitative measure of evidentiary weight, then verbatim representations of the examiner’s testimony may have played a greater role in judgments. Future studies should more closely examine how examiner testimony is encoded and the relative impact of gist and verbatim representations of examiner testimony on evidential combinations and jury verdicts.

Our basic finding about the lack of impact of different ways of framing an examiner’s match conclusion should be viewed in conjunction with our finding that people generally view fingerprint identifications as reliable and believe 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 to be a strong statement of individualization that needs no bolstering. The bolstering language appears to have just reaffirmed what layperson already thought 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 pre-existing 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 lay jurors’ consideration of fingerprint evidence. Our studies indicate that the public has unrealistically high expectations about the error rates associated with fingerprint identifications, but also indicate that laypersons are sensitive to information on error rates.

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 (2012) 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, 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 discussions. Our goal was to examine the marginal effects of fingerprint evidence, compared to the control case, on lay persons 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, that may in turn color the way in which evidence is interpreted and assimilated (e.g., Mazzocco, Alicke & Davis, 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 further study. Our results suggest that the precise terminology used by fingerprint examiners to convey a positive match conclusion may have little impact on the weight given to that identification, even when that terminology makes exorbitant claims about the certainty of the match. Thus, the focus by the NRC and the NIST/NIJ task force on forensic standards on specific match terminology may be misplaced. Most importantly, 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 may significantly affect the weight given to an examiner’s testimony. Given the indisputable fact that fingerprint examinations are not infallible and another person 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 pre-existing beliefs about fingerprint evidence and portray examiner testimony in an overly positive light. As one fingerprint examiner put it recently, “when 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, p. 8).

Greater attention should be paid to the problem of how to express conclusions 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 an expert states an ultimate conclusion, and the confidence with which that conclusion is stated, may be much less important than the background information provided by the expert and the degree to which that information confirms or contradicts assumptions that jurors already hold about the reliability and probative value of the expert’s method and opinions.
References


Eldridge, H. (March-April, 2012). “I am 100% certain of my conclusion.” (But should the jury be certain?) *Evidence Technology Magazine*, 8.


International Association for Identification. (July 16, 2010). IAI Resolution 2010-18.


U.S. v. Baines, 573 F.3d 979 (10th Cir. 2009).

State v. Quintana, 103 P.3d 168 (Utah 2004).
Table 1

Examples of Fingerprint Examiner Terminology

<table>
<thead>
<tr>
<th>Simple Positive Match</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A latent fingerprint found at the scene was individualized as the left thumb of the defendant.</td>
<td></td>
</tr>
<tr>
<td>2. The latent fingerprint found at the scene matched the left thumb print on the ink card labeled as taken from the defendant.</td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>4. I am confident that the defendant was the source of the latent fingerprint found at the crime scene.</td>
<td></td>
</tr>
<tr>
<td>5. I concluded that it is very likely that the defendant was the source of the latent fingerprint found at the crime scene.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bolstered Positive Match</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Qualified or Inconclusive Match</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>14. I cannot exclude the defendant as the source of the latent fingerprint found at the scene.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15. The friction ridge impression excluded the left thumb of the defendant and originated from a different source.</td>
<td></td>
</tr>
</tbody>
</table>
**Note.** 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, U.S. v. Baxter et al., F-5928-01 (D.C. Sup. Ct. June 16, 2003); see also Cole, 2007, p. 127); No. 3: based on an F.B.I. examiner’s testimony that a print “was made by the same individual” (Trial Testimony at 105, U.S. v. Latrell Gilchrist, No. F-2158-01 (D.C. Sup. Ct. March 13, 2003); see also Cole, 2007, p. 126); Nos. 4 and 5: this type of source attribution testimony is apparently common (see Cole, 2007, p. 473); the language of No. 4 was based on testimony of an F.B.I. analyst that “I am confident in the conclusion that I rendered” (Hearing Transcript at 124, U.S. 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 case fingerprint case and the sample language has been proposed for fingerprint cases (Cole, 2011, p. 472); No. 9: this language comes directly from current SWGFAST guidelines (SWGFAST, 2011); No. 10: this language comes directly from testimony of a former F.B.I. supervisor (Hearing Transcript at 50, U.S. v. Baines, No. 06-CR-1797 (D.N.M. April 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.
### Table 2
Study 1: Dependent Measure Means by Testimony Condition

<table>
<thead>
<tr>
<th>Fingerprint Evidence Condition</th>
<th>Likelihood Defendant Robber (1-7 scale)</th>
<th>Confidence in Likelihood Judgment (1-5 scale)</th>
<th>Probability Defendant Left Fingerprints at Scene (0-100 scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Control</td>
<td>3.16&lt;sub&gt;a&lt;/sub&gt; (1.14)</td>
<td>3.42&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.93&lt;/sub&gt;)</td>
<td>---</td>
</tr>
<tr>
<td>Simple Positive Match</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.28&lt;sub&gt;b&lt;/sub&gt; (1.35)</td>
<td>3.19&lt;sub&gt;b&lt;/sub&gt; (&lt;sub&gt;.88&lt;/sub&gt;)</td>
<td>63.67&lt;sub&gt;a&lt;/sub&gt; (23.09)</td>
</tr>
<tr>
<td>2</td>
<td>4.54&lt;sub&gt;b&lt;/sub&gt; (1.22)</td>
<td>3.52&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.84&lt;/sub&gt;)</td>
<td>63.84&lt;sub&gt;b&lt;/sub&gt; (27.97)</td>
</tr>
<tr>
<td>3</td>
<td>4.34&lt;sub&gt;b&lt;/sub&gt; (1.22)</td>
<td>3.25&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.86&lt;/sub&gt;)</td>
<td>62.48&lt;sub&gt;b&lt;/sub&gt; (23.30)</td>
</tr>
<tr>
<td>4</td>
<td>4.40&lt;sub&gt;b&lt;/sub&gt; (1.38)</td>
<td>3.22&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.99&lt;/sub&gt;)</td>
<td>64.99&lt;sub&gt;,ab&lt;/sub&gt; (24.52)</td>
</tr>
<tr>
<td>5</td>
<td>4.36&lt;sub&gt;b&lt;/sub&gt; (1.22)</td>
<td>3.24&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.99&lt;/sub&gt;)</td>
<td>62.40&lt;sub&gt;,ab&lt;/sub&gt; (24.52)</td>
</tr>
<tr>
<td>Bolstered Positive Match</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4.52&lt;sub&gt;b&lt;/sub&gt; (1.34)</td>
<td>3.30&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.97&lt;/sub&gt;)</td>
<td>64.92&lt;sub&gt;a&lt;/sub&gt; (25.78)</td>
</tr>
<tr>
<td>7</td>
<td>4.40&lt;sub&gt;b&lt;/sub&gt; (1.40)</td>
<td>2.99&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.98&lt;/sub&gt;)</td>
<td>63.71&lt;sub&gt;a&lt;/sub&gt; (23.98)</td>
</tr>
<tr>
<td>8</td>
<td>4.26&lt;sub&gt;b&lt;/sub&gt; (1.27)</td>
<td>3.04&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.93&lt;/sub&gt;)</td>
<td>60.28&lt;sub&gt;,b&lt;/sub&gt; (24.08)</td>
</tr>
<tr>
<td>9</td>
<td>4.36&lt;sub&gt;b&lt;/sub&gt; (1.42)</td>
<td>3.21&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.95&lt;/sub&gt;)</td>
<td>66.03&lt;sub&gt;,b&lt;/sub&gt; (26.29)</td>
</tr>
<tr>
<td>10</td>
<td>4.57&lt;sub&gt;b&lt;/sub&gt; (1.34)</td>
<td>3.24&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.92&lt;/sub&gt;)</td>
<td>66.53&lt;sub&gt;,b&lt;/sub&gt; (23.79)</td>
</tr>
<tr>
<td>11</td>
<td>4.56&lt;sub&gt;b&lt;/sub&gt; (1.43)</td>
<td>3.39&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.80&lt;/sub&gt;)</td>
<td>70.04&lt;sub&gt;,b&lt;/sub&gt; (25.47)</td>
</tr>
<tr>
<td>Qualified or Inconclusive Match</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.03&lt;sub&gt;a&lt;/sub&gt; (1.13)</td>
<td>3.15&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;1.01&lt;/sub&gt;)</td>
<td>37.66&lt;sub&gt;,d&lt;/sub&gt; (23.17)</td>
</tr>
<tr>
<td>13</td>
<td>3.50&lt;sub&gt;a&lt;/sub&gt; (1.18)</td>
<td>3.09&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;1.00&lt;/sub&gt;)</td>
<td>48.71&lt;sub&gt;,b&lt;/sub&gt; (25.57)</td>
</tr>
<tr>
<td>14</td>
<td>3.56&lt;sub&gt;a&lt;/sub&gt; (1.02)</td>
<td>3.01&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.91&lt;/sub&gt;)</td>
<td>49.56&lt;sub&gt;,b&lt;/sub&gt; (23.46)</td>
</tr>
<tr>
<td>Exclusion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3.06&lt;sub&gt;a&lt;/sub&gt; (1.30)</td>
<td>3.15&lt;sub&gt;a&lt;/sub&gt; (&lt;sub&gt;.96&lt;/sub&gt;)</td>
<td>41.96&lt;sub&gt;,cd&lt;/sub&gt; (28.06)</td>
</tr>
</tbody>
</table>

*Note.* Condition number corresponds to testimony number in Table 1. Means in the same column not sharing the same subscript are significantly different at $p < .05$. Tests assume equal variances, and the tests were adjusted for multiple pairwise comparisons using the Bonferroni correction.
Table 3
Study 2: Dependent Measure Means by Testimony Condition

<table>
<thead>
<tr>
<th>Fingerprint Evidence Condition</th>
<th>Likelihood Defendant Robber (1-7 scale)</th>
<th>Confidence in Likelihood Judgment (1-5 scale)</th>
<th>Probability Defendant Left Fingerprints at Scene (0-100 scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Control</td>
<td>3.30 (.94)</td>
<td>3.15 (.95)</td>
<td>---</td>
</tr>
<tr>
<td>Simple Individualization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Method Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDD</td>
<td>5.41 (1.24)</td>
<td>3.56 (.87)</td>
<td>75.41 (22.38)</td>
</tr>
<tr>
<td>EDC</td>
<td>5.38 (1.10)</td>
<td>3.52 (1.04)</td>
<td>74.03 (22.16)</td>
</tr>
<tr>
<td>EAD</td>
<td>4.93 (1.03)</td>
<td>3.33 (.73)</td>
<td>68.34 (21.57)</td>
</tr>
<tr>
<td>EAC</td>
<td>4.66 (1.28)</td>
<td>3.34 (.79)</td>
<td>63.02 (20.63)</td>
</tr>
<tr>
<td>Method Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDD</td>
<td>5.44 (1.23)</td>
<td>3.71 (.81)</td>
<td>77.27 (20.96)</td>
</tr>
<tr>
<td>EDC</td>
<td>5.37 (1.12)</td>
<td>3.65 (.80)</td>
<td>71.12 (27.39)</td>
</tr>
<tr>
<td>EAD</td>
<td>5.20 (1.13)</td>
<td>3.41 (.89)</td>
<td>74.71 (22.26)</td>
</tr>
<tr>
<td>EAC</td>
<td>5.11 (1.23)</td>
<td>3.42 (.76)</td>
<td>69.34 (26.69)</td>
</tr>
<tr>
<td>Bolstered Individualization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Method Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDD</td>
<td>5.51 (1.03)</td>
<td>3.66 (.67)</td>
<td>74.27 (23.87)</td>
</tr>
<tr>
<td>EDC</td>
<td>5.48 (1.30)</td>
<td>3.85 (.76)</td>
<td>78.57 (20.06)</td>
</tr>
<tr>
<td>EAD</td>
<td>4.93 (1.37)</td>
<td>3.48 (.78)</td>
<td>65.95 (27.28)</td>
</tr>
<tr>
<td>EAC</td>
<td>4.59 (1.12)</td>
<td>3.39 (.92)</td>
<td>63.34 (23.31)</td>
</tr>
<tr>
<td>Method Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDD</td>
<td>5.20 (1.35)</td>
<td>3.70 (.83)</td>
<td>76.56 (26.16)</td>
</tr>
<tr>
<td>EDC</td>
<td>5.51 (1.36)</td>
<td>3.85 (.82)</td>
<td>81.56 (19.59)</td>
</tr>
<tr>
<td>EAD</td>
<td>5.21 (1.17)</td>
<td>3.63 (.71)</td>
<td>74.38 (22.69)</td>
</tr>
<tr>
<td>EAC</td>
<td>4.68 (1.35)</td>
<td>3.46 (.87)</td>
<td>65.61 (29.35)</td>
</tr>
</tbody>
</table>

Note. EDD = error risk discounted on direct; EDC = error risk discounted on cross; EAD = error risk acknowledged on direct; EAC = error risk acknowledged on cross.