

Certainty & Uncertainty in Reporting Fingerprint Evidence

Joseph B. Kadane & Jonathan J. Koehler

Abstract: Everyone knows that fingerprint evidence can be extremely incriminating. What is less clear is whether the way that a fingerprint examiner describes that evidence influences the weight lay jurors assign to it. This essay describes an experiment testing how lay people respond to different presentations of fingerprint evidence in a hypothetical criminal case. We find that people attach more weight to the evidence when the fingerprint examiner indicates that he believes or knows that the defendant is the source of the print. When the examiner offers a weaker, but more scientifically justifiable, conclusion, the evidence is given less weight. However, people do not value the evidence any more or less when the examiner uses very strong language to indicate that the defendant is the source of the print versus weaker source identification language. We also find that cross-examination designed to highlight weaknesses in the fingerprint evidence has no impact regardless of which type of conclusion the examiner offers. We conclude by considering implications for ongoing reform efforts.

JOSEPH B. KADANE, a Fellow of the American Academy since 2010, is the Leonard J. Savage University Professor of Statistics and Social Sciences, Emeritus, at Carnegie Mellon University. His most recent book is *Pragmatics of Uncertainty* (2017).

JONATHAN J. KOEHLER is the Beatrice Kuhn Professor of Law at the Northwestern University Pritzker School of Law. He conducts research related to decision-making in the courtroom. His work has recently been published in such journals as *Journal of Criminal Law and Criminology*, *Jurimetrics Journal*, and *Psychology, Public Policy, and Law*.

The study of fingerprints began in a serious way with Francis Galton's book *Finger Prints* in 1892.¹ For more than a century, fingerprint results were treated by the forensic science community (and the courts) as infallible, or nearly so. In 1985, an authoritative FBI manual stated, "of all the methods of identification, fingerprinting alone has proved to be both infallible and feasible."² In a 2003 segment on the television news program *60 Minutes*, the head of the FBI's fingerprint unit said that the probability of error in fingerprint analysis is 0 percent, and that all analysts are and should be 100 percent certain of the identifications that they offer in court.³ Such hyperbole is unscientific and unsustainable. As it turned out, just a few months after this program aired, the FBI was forced to admit that its top fingerprint examiners matched a print to the wrong person in the investigation of the 2004 Madrid train bombings, one of the highest profile fingerprint cases in history.⁴

© 2018 by Joseph B. Kadane & Jonathan J. Koehler
doi:10.1162/DAED_a_00524

When the National Academy of Sciences (NAS) completed its comprehensive review of many of the non-DNA forensic sciences (including fingerprint evidence) in 2009, the results were shocking.⁵ The NAS found that many of the most basic forensic science claims had not been validated by empirical research. In response, federal agencies and forensic science professional organizations began working in earnest to, among other things, modify the ways in which forensic scientists present evidence in court. Simple and obvious reforms such as eliminating references to “100 percent certain” identifications and “0 percent risk of error” have already taken hold. However, forensic science reformers have been largely flying blind on the question of which specific words should replace the exaggerated conclusions that forensic scientists often provide in their courtroom testimony.⁶

Fingerprint evidence has been admitted in U.S. courts as proof of identity in criminal cases for more than one hundred years. Evidence that an unknown print recovered from a crime scene (a so-called latent print) matches a known print from a suspect or other individual is rarely challenged in court and is widely regarded by the public as conclusive proof that the person whose fingerprint matched is the source of the latent print in question. This is the case even when the match is to latent prints that are partial, smudged, or otherwise of low quality, although all of these features increase both the difficulty of declaring a definite match and the likelihood of error.⁷

Fingerprint evidence has long been a powerful tool for criminal investigators and prosecutors. A latent print found on, say, a gun recovered from a shooting scene not only helps identify a person of interest for police in the early stages of an investigation, but it also may be the single most powerful proof of a defendant’s guilt of-

ferred at trial. This essay looks at the use of fingerprint evidence as a tool to persuade judges and jurors at trial or, more commonly, to persuade a criminal defendant to accept a plea bargain rather than risk a seemingly certain guilty verdict. This essay, however, concerns itself only with the presentation of fingerprint evidence at trials, asking how the way a fingerprint examiner testifies about his or her results affects the weight that factfinders are likely to assign to the evidence.

In the typical case involving fingerprint evidence, a trained examiner compares one or more latent prints with various known or exemplar prints using a high-powered microscope. This process is often preceded by an automated search through a local, state, or national database. The national database includes fingerprints from approximately 120 million people. The computer search narrows the list of candidate prints and orders them so that the most likely matches appear at the top of the list. The examiner then proceeds to make pairwise comparisons between the latent and candidate prints. The end result of the pairwise comparison process (known as ACE-V)⁸ is usually one of four conclusions: *identification* (the prints come from the same source), *exclusion* (the prints come from different sources), *inconclusive*, or *unsuitable for comparison*.

Although the ACE-V process is subjective, fingerprint examiners have historically claimed that their identifications are 100 percent certain, and that there is virtually no chance that an error has occurred.⁹ The precise meaning of the word “identification” may, however, vary depending on idiosyncratic definitions and usage by various parties.¹⁰ While the Humpty-Dumpty dictum may appeal to some testifying forensic scientists (“when I use a word . . . it means just what I choose it to mean”), it is unjustified in courts of law where the interpretation of an unfamiliar or techni-

cal phrase may be the difference between freedom and incarceration for a criminal defendant.¹¹

There is some basis in the broader forensic science literature for suggesting that the weight that jurors assign to fingerprint evidence will depend, in part, on the way that fingerprint examiners describe their conclusions. Psychological studies show that the way DNA results are framed affects the value that people accord to reported matches.¹² Focusing on microscopic hair results, psychologists Dawn McQuiston-Surrett and Michael Saks found that the way hair-evidence matches are described affects mock jurors' assessments of the probability that the person whose hair is said to match is the source of the unknown hair.¹³ The mock jurors in their experiments assigned higher probabilities of source identification when hair evidence was described as either a "match" or "similar in all microscopic characteristics," and lower probabilities when the forensic expert estimated the number of other people in the city who would also match.

However, these studies did not find that mock jurors' judgments varied as a function of whether forensic experts went further and volunteered their own opinion that the person whose hair was said to match was the source of the hair. Similarly, an experiment conducted by legal scholar Brandon Garrett and psychologist and lawyer Gregory Mitchell in the context of short written cases that involved fingerprint evidence found that "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."¹⁴ Participants in their study were no more impressed with the fingerprint evidence when the fingerprint examiner said that it was "a practical impossibility" that someone other than the defendant was the source of the

latent print or the examiner simply said that the defendant "matched" the latent print.¹⁵ These investigators concluded that it really did not matter how an examiner framed a match conclusion because factfinders give "considerable weight" to fingerprint match evidence in all forms.¹⁶

It appears, then, that the way forensic science testimony is presented by experts will matter in some circumstances but not others. When a defendant is a member of a group of those who might be the source of an incriminating piece of evidence, testimony that fails to point out the existence of others in the set who might be the source is viewed as stronger than testimony that expressly notes that the defendant is one of a group of people who might be the source. But when a forensic scientist indicates in some fashion his or her belief that the defendant is the source of an incriminating piece of evidence, it is not clear that such add-on comments have an extra impact on factfinders.

Thanks in large part to the 2009 National Academy of Sciences report on the state of the non-DNA forensic sciences, efforts are underway to standardize and reform many forensic science practices, including the way results and conclusions are reported in court.¹⁷ These reform efforts have thus far proceeded with little or no guidance from empirical studies. Consequently, there is a risk that proposed changes in the conclusory language used by forensic scientists will have no impact – or perhaps even an unintended impact – on police, legal decision makers, and others who rely on forensic evidence. Our essay reports on an experiment that addresses this concern. The experiment examines how people interpret and use different verbal formulations of conclusions reached by fingerprint examiners.

Six hundred jury-eligible citizens (U.S. citizens, at least eighteen years old, with

Joseph B.
Kadane &
Jonathan J.
Koehler

no felony convictions) were paid to answer an online questionnaire about a hypothetical legal case that included fingerprint evidence. Our mock jurors (“jurors”) covered a broad and representative cross-section of the jury-eligible population in terms of education level (19.5 percent high school diploma or less, 14.5 percent graduate degrees), ethnicity (7.7 percent African American), and gender (58 percent women).

Jurors were presented with the following scenario:

In a recent legal case, Mr. Richard Johnson was charged with robbing a convenience store. Although the perpetrator of the crime wore a hood that covered his face, Mr. Johnson became a suspect when the store owner told police that he thought the perpetrator sounded very much like one of his frequent customers, Mr. Richard Johnson. The store owner also told police that the perpetrator reached into the opened cash register with his bare hand and lifted one of the trays. When a police fingerprint examiner examined the cash register and its inside trays for fingerprints, he found 19 prints that were suitable for comparison purposes. The fingerprint examiner eliminated Mr. Johnson as a potential source of 18 of those prints. However, the fingerprint examiner was not able to eliminate Mr. Johnson as a possible contributor of one of the prints that was found on the cash register tray. At Mr. Johnson’s robbery trial, the fingerprint examiner was called to testify for the prosecution. After the fingerprint examiner discussed his credentials, experience, and methods, the following exchange took place between the prosecutor (P) and the fingerprint examiner (FE):

P: Now you said that you recovered 19 fingerprints from the cash register, is that correct?

FE: Yes, there were 19 prints that had enough detail that I could compare them to known exemplars.

P: What is a known exemplar?

FE: It’s a reference print – a print whose source is known. We compare the prints that we recover on objects from a crime scene with various known exemplars. So in this case, I had known exemplars from Mr. Johnson, the employees of the convenience store, and some other people. And I compared the prints that were on the cash register and cash register components with the known exemplars.

P: OK, and what were your findings with respect to the prints that were on the cash register and the known exemplar print provided by the defendant in this case, Mr. Richard Johnson?

FE: Well, first, I was able to exclude Mr. Johnson as a possible contributor of 18 of the 19 latent prints that were on the cash register or its various components. In other words, none of those 18 prints were made by Mr. Johnson. However, I was not able to exclude Mr. Johnson as a possible contributor of the 19th print. This 19th print was taken from the cash register tray.

P: And so your bottom line conclusion is what?

At this point, jurors received one of six different single-sentence conclusions from the fingerprint examiner. In all cases, the conclusion that jurors saw was preceded by the words “My bottom line conclusion is that . . .” The six conclusory statements were as follows:

1. “. . . I cannot exclude the defendant, Mr. Johnson, as a possible contributor of that print.”
2. “. . . the likelihood of observing this amount of correspondence when two impressions are made by different sources is considered extremely low.”
3. “. . . in my opinion, the defendant, Mr. Johnson, is the source of that print.”

4. "... in my opinion, the defendant, Mr. Johnson, is the source of that print to a reasonable degree of scientific certainty."
5. "... I was able to effect an individualization on that latent print to the defendant, Mr. Johnson."
6. "... I was able to effect an individualization on that latent print to the defendant, Mr. Johnson, to the exclusion of all other possible sources in the world."

The first conclusion ("I cannot exclude") is widely recognized as a scientifically accurate and defensible (albeit conservative) way to describe the results of a match between a known and unknown print.¹⁸ If a known print from a suspect appears to share a common set of characteristics with an unknown print recovered from a crime scene and there are no other explainable inconsistencies, it follows as a matter of logic that an examiner would be justified in concluding that the suspect cannot be excluded as a possible contributor of the unknown print. However, a significant shortcoming of this conclusion is that it does not specify the size of the nonexcluded class of individuals.

The second conclusion reflects the language that has been recommended by the U.S. Army.¹⁹ It is essentially a statement that the false positive error rate is "extremely low." Because this conclusion does not specify what is meant by "extremely low," it is hard for anyone to know how much weight to assign to this evidence.

The third conclusion may be defensible under the Federal Rules of Evidence, but as a purely scientific matter, it is also less defensible than the first conclusion because the examiner is making an inferential leap from evidence indicating that a suspect may be the source of a print to a personal conclusion that the suspect is, in fact, the source of that print.²⁰ Even if the available science gave the examiner good reason to

believe that the class of people who might be the source of the print is very small, a source claim involves a degree of speculation and guesswork that extends beyond what the science can show.²¹

The fourth conclusion suffers from the same problem as the third, but is even more objectionable because it appends the impressive-sounding but scientifically meaningless phrase "to a reasonable degree of scientific certainty" to the examiner's personal opinion. The result may be inflated confidence or simply greater variability in understanding the level of confidence it is intended to convey.²²

The fifth conclusion, which has long been favored by the FBI, goes further by replacing the "opinion" language in conclusions three and four with "individualization" language.²³ Use of this language might give the misleading impression that the science itself has unequivocally identified the source of the print.²⁴

The sixth conclusion is even stronger than the fifth because it expressly states that the individualization has excluded all other possible sources in the world.

In sum, the first conclusion is the least objectionable from the standpoint of science and logic, though it is far from satisfying. The second conclusion is problematic because it does not explain what an "extremely low" chance of a coincidental match means. Conclusions three through six all involve a questionable scientific leap of faith in moving from the absence of proof that two prints come from different sources to a finding that the two prints must have come from a common source.

Returning to the experiment, after the fingerprint examiner stated his or her conclusion, the prosecutor repeated the examiner's conclusion verbatim, as many prosecutors do to ensure that jurors don't miss the examiner's conclusion. The examiner confirmed that this was indeed his conclusion.

Joseph B.
Kadane &
Jonathan J.
Koehler

Half of the jurors (Conditions 1 – 6) read a cross-examination that was tailored to challenge the specific conclusion used by the fingerprint examiner. For example, when the fingerprint examiner said that he was able to “effect an individualization” on that latent print to the defendant, Mr. Johnson, “to the exclusion of all other possible sources in the world” (Condition 6), the cross-examiner elicited a confession from the witness that he has not actually examined the prints of all other people in the world. Likewise, when the examiner says, “in my opinion, the defendant, Mr. Johnson, is the source of that print” (Condition 3), the cross-examiner elicits a confession from the expert witness that he is not claiming that he absolutely positively knows that the print came from Mr. Johnson’s finger, to the exclusion of all other possible sources in the world. The other half of the jurors were assigned to a no-cross examination condition (Conditions 7 – 12). Table 1 summarizes the twelve conditions. Whether or not they read a cross-examination, jurors in all conditions answered the same set of questions about the case.

We asked jurors four “source” questions about the value of the fingerprint evidence for the proposition that the fingerprint belonged to the defendant, Mr. Johnson. Questions 1 – 3 and 5 used a scale that ranged from 1 (not at all) to 7 (extremely):

1. How strong would you say the fingerprint evidence is with respect to the prosecutor’s claim that the fingerprint on the cash register tray belongs to Mr. Johnson (the defendant)?
2. How convincing would you say the fingerprint evidence is with respect to the prosecutor’s claim that the fingerprint on the cash register tray belongs to Mr. Johnson (the defendant)?
3. How confident are you that the fingerprint on the cash register tray was left by the defendant?

4. What would you say is the probability that the fingerprint on the cash register tray belongs to the defendant? (Please provide a number between 0% and 100%.)

Next, we asked two “guilt” questions about jurors’ beliefs that the defendant, Mr. Johnson, committed the robbery:

5. How confident are you that the fingerprint on the cash register tray was left by the defendant during the course of the convenience store robbery?
6. What would you say is the probability that the defendant robbed the convenience store? (Please provide a number between 0% and 100%.)

The answers participants provided to the four source and two guilt questions were all highly correlated with one another ($0.69 < r\text{'s} < 0.88$). We therefore created an aggregated “strength of evidence” index for each participant that gave equal weight to the six questions asked.

The next task was to compare the conditions with cross-examination to those without. To do so, we combined the indices for participants in Conditions 1 to 6 into an index with cross-examination. Similarly, we combined the indices for participants in Conditions 7 to 12 into an index without cross-examination. The data indicated that there was no effect for cross-examination. If anything, our subjects found the evidence without cross-examination slightly more plausible than the evidence with cross-examination, but the difference is so slight that we can ignore it. This permits us to aggregate Conditions 1 and 7, 2 and 8, and so on, giving us only six conditions (distinguished by the wording used by the fingerprint examiner). When we refer in the rest of this essay to Condition 1, for example, what we mean is the aggregation of Conditions 1 and 7 in Table 1; the same is true of all Conditions 1 – 6.

Table 1
Twelve Conditions

Joseph B.
Kadane &
Jonathan J.
Koehler

Expert Testimony	Condition	
	Cross-Examination	No Cross-Examination
Cannot exclude Mr. Johnson	1	7
The likelihood of observing this amount of correspondence when two impressions are made by different sources is considered extremely low	2	8
Mr. Johnson is the source	3	9
Mr. Johnson is the source to a reasonable degree of scientific certainty	4	10
I effected an individualization on that print to Mr. Johnson	5	11
I effected an individualization on that print to Mr. Johnson to the exclusion of all possible other sources in the world	6	12

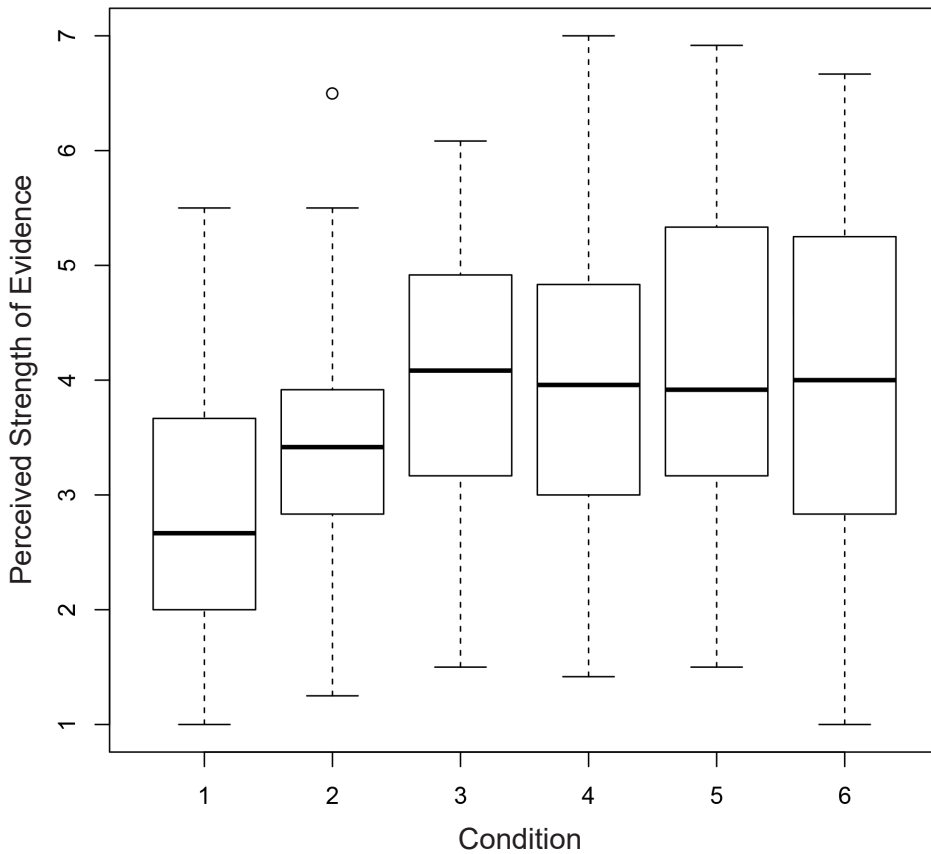
Whether another style of cross-examination would show a larger effect is not addressed by our data.

Our primary analysis focuses on how differences in the language that the fingerprint examiner used to describe the match evidence affected subjects' judgments about the strength of the evidence. We do this by comparing the evidence strength index scores across the six fingerprint examiner conditions. In Figure 1, arrayed along the horizontal axis are the conditions, and the vertical axis is the index of strength of evidence. Figure 1 shows that the language used by the fingerprint examiner in Condition 1 ("cannot exclude") was the least impactful way of reporting the fingerprint evidence, followed by Condition 2 ("the likelihood of observing this amount of correspondence . . . is considered extremely low"). The language used to describe fingerprint evidence in Conditions

3 – 6 was more impactful than that of Conditions 1 and 2, and differed little by condition. To the extent our results generalize to actual trials, we see the importance of how forensic scientists present their testimony, and the need to ensure that the language a forensic scientist uses fairly reflects the evidentiary implications of the reported evidence.

The language used by the fingerprint examiner in the six conditions was designed to vary in the certitude with which the examiner provided his conclusion. For example, an examiner who says that he has "effected an individualization on that print to Mr. Johnson to the exclusion of all possible other sources in the world" (Condition 6) appears to be expressing much greater certainty in his conclusion than an examiner who simply says that Mr. Johnson cannot be excluded as a possible contributor of the

Figure 1
Perceived Strength of Evidence by Condition



Note: Condition 1 = cannot exclude; Condition 2 = extremely low likelihood of such correspondence by different sources; Condition 3 = source of the print; Condition 4 = source of the print to a reasonable degree of scientific certainty; Condition 5 = individualization; Condition 6 = individualization to the exclusion of all other possible sources in the world. The rectangular boxes above each condition capture the interquartile range, and the bold horizontal line within each box shows the median for that condition.

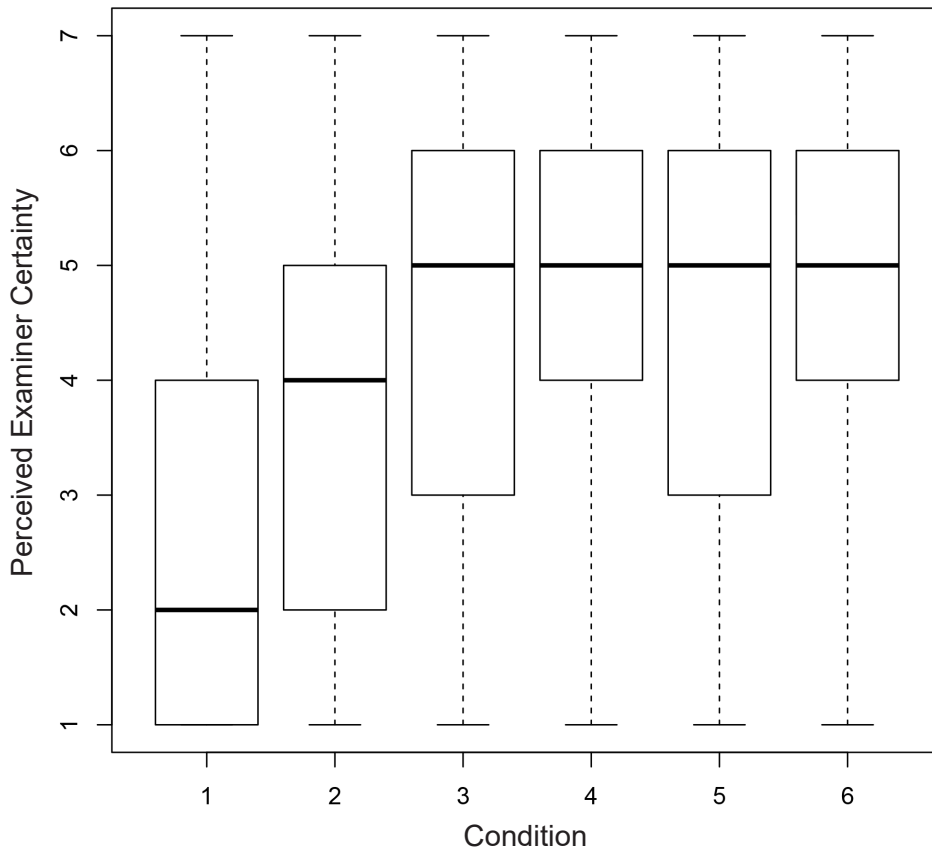
same print (Condition 1). We checked this assumption by asking jurors the following question: “How certain was the expert that the defendant was the source of the fingerprint on the cash register tray?” The distribution of jurors’ answers across the six conditions is plotted in Figure 2. Here the vertical axis is a scale of certainty (1 = low, 7 = high). As expected, the data show that people believe that the fingerprint examin-

er is least certain in Condition 1, followed by Condition 2. Further, our jurors believed that the examiner was more certain of his conclusion in Conditions 3 – 6 than in Conditions 1 and 2. It is notable that the medians for Conditions 3 – 6 are identical.

We also asked our participants whether the uncertainty expressed by the fingerprint examiner mattered: “How much does it matter in a case like this whether the ex-

Figure 2
Perceived Examiner Certainty by Condition

Joseph B.
Kadane &
Jonathan J.
Koehler



pert witness is certain about his conclusions (rather than expressing uncertainty)?” The degree of certainty expressed by the fingerprint examiner mattered a great deal to our jurors in all conditions (median ratings of 6 or 7 out of 7).

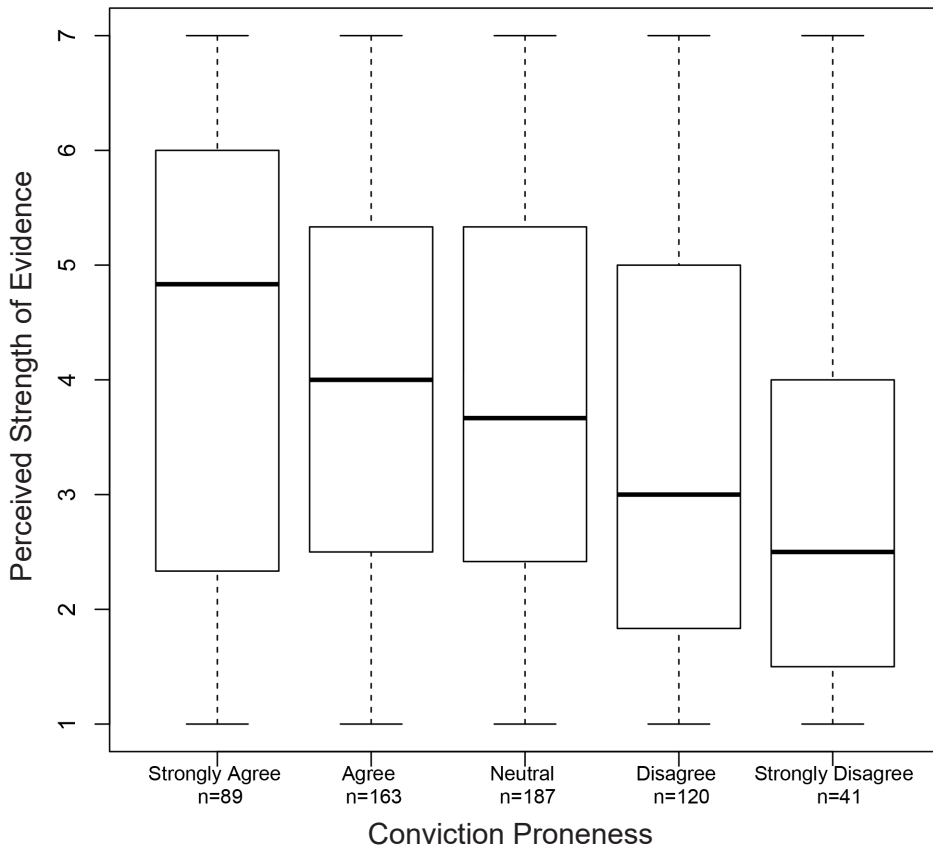
In addition to seeking judgments about the weight the fingerprint evidence deserved, we sought demographic and opinion information from our respondents. We found that men, African Americans, and those with graduate degrees are somewhat more skeptical of fingerprint evidence than others. Jury service and law enforcement service (self or relative), political leanings (liberal or conservative),

and frequency of watching CSI or similar television shows had no effect on indexed responses. However, we did find a strong relationship between index scores and responses to the item below:

Our criminal justice system should be less concerned about protecting the rights of the people charged with crimes and more concerned about convicting the guilty. (Please select only one)

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree.

Figure 3
Perceived Strength of Evidence by Conviction Proneness



Note: Participants who marked Strongly Agree or Agree to the conviction proneness question indicated a relatively strong concern for convicting the guilty, and participants who marked Disagree or Strongly Disagree indicated a relatively strong concern for protecting defendants' rights.

Figure 3 shows that respondents who thought we should be more concerned about convicting the guilty (as reflected by agreement or strong agreement with the statement above) tended to assign more weight to the fingerprint evidence across conditions. It is not surprising that this should be so. Perhaps it is further evidence that “we see things not as they are, but as we are.”

To summarize, participants in our study attached more weight to the fingerprint

evidence in the four conditions (3–6) in which the examiner indicated in some manner that he or she believes or knows that the defendant is the source of the print than when the examiner offered a weaker, but more scientifically justifiable, conclusion (Conditions 1 and 2). The phrases commonly used to bolster source opinion and individualization claims (“to a reasonable degree of scientific certainty” and “to the exclusion of all other people in the world,” respectively) had no apprecia-

ble effect on the judgments of our jurors. A simple cross-examination that was tailored to highlight weaknesses in the fingerprint evidence in each condition likewise had no impact regardless of which type of conclusion the examiner offered. Gender, race/ethnicity, education, political leaning, jury service, and law enforcement service (their own or that of a relative) produced only minor effects, though we did find that people who indicated that our criminal justice system should be more concerned with convicting the guilty tended to assign greater weight to the fingerprint evidence.

Fingerprint analysts may find that a latent print and a known print share certain characteristics. However, at the present time, they have no scientific way to estimate the number of people in a given population whose fingerprints are likely to share those characteristics.²⁵ In this respect, fingerprint analysis differs from DNA analysis because only the latter has systematically documented the frequency of the relevant characteristics among various populations. Consequently, there is insufficient scientific justification for a claim that the person whose fingerprint matches that of a latent print recovered from a crime scene must be the source of that print. For this reason, we believe the source and individualization statements in some of our conditions overstate the strength of the evidence.

On June 3, 2016, the Department of Justice proposed “uniform language for testimony and reports for the forensic latent print discipline.”²⁶ This proposal included approval for a finding of identification, but barred “to the absolute exclusion of all others” and “a zero error rate or ... infallible.” Our results show that the proposed limitations are unlikely to affect how lay persons, such as judges and jurors, understand latent print testimony. However, we

did find that when the identification language is abandoned in favor of the weaker, but more scientifically justifiable, “cannot be excluded” conclusion, people attached less weight to the fingerprint evidence. If future researchers are able to identify the frequency with which various print features arise in the population, then perhaps the cannot-be-excluded conclusion could be modified to include an estimate of the number of people who could be the source of the latent print in question. If that group is sufficiently small, it seems likely that people will attach more weight to fingerprint evidence that is presented with the further empirically justified information attached.

The President’s Council of Advisors on Science and Technology (PCAST) reported in 2016 that

Based largely on two recent appropriately designed black-box studies, PCAST finds that latent fingerprint analysis is a foundationally valid subjective methodology – albeit with a false positive rate that is substantial and is likely to be higher than expected by many jurors based on longstanding claims about the infallibility of fingerprint analysis. Conclusions of a proposed identification may be scientifically valid, provided that they are accompanied by accurate information about limitations on the reliability of the conclusion – specifically, that (1) only two properly designed studies of the foundational validity and accuracy of latent fingerprint analysis have been conducted, (2) these studies found false positive rates that could be as high as 1 error in 306 cases in one study and 1 error in 18 cases in the other, and (3) because the examiners were aware they were being tested, the actual false positive rate in casework may be higher.²⁷

The two studies referred to in the PCAST report come from Noblis researcher Bradford T. Ulery and colleagues in 2011 and 2012.²⁸ We are less impressed with these

Joseph B.
Kadane &
Jonathan J.
Koehler

studies than was PCAST for several reasons. First, the subjects were volunteers who knew they were being tested. Second, the studies were paid for by the FBI (an interested party) and some of the authors worked for the FBI. Third, the proportion of judgments of identification and exclusion varied widely, suggesting that some examiners were very cautious, perhaps more cautious than they would be in casework. This reduces the credibility of the false positive and false negative rates found in these studies. Nonetheless, such studies are useful for comparing groups of fingerprint examiners and for comparing the difficulty of different types of fingerprint assessment tasks.

Although our empirical study – which obviously was not designed to measure what the Ulery studies measured – does not share these shortcomings, our results should likewise be interpreted with caution. It is a single study, conducted online with individual participants who had no opportunity to test their reactions by comparing them with those of others, and the precise wording of our stimuli and questions may have influenced the answers provided.²⁹ Further, because we used just one scenario and a single forensic technique (fingerprinting), it is difficult to say how well our results generalize either to other fingerprint scenarios or other forensic science methods.

Having said that, our results appear to reinforce and extend the observation by McQuiston-Surrett and Saks, and Garrett and Mitchell that lay people may not be sensitive to distinctions between stronger and weaker conclusions that an expert draws about forensic matching evidence once the expert has declared a match or words to that effect.³⁰ In those studies, the judgment made by mock jurors did not vary as a function of whether the forensic expert provided an opinion about whether matching hairs came from the same person (McQuiston-Surrett and Saks) or whether matching prints were described as not possibly be-

longing to anyone other than the defendant (Garrett and Mitchell). Likewise, our mock jurors did not draw distinctions among different “source” claims, including those that were designed to impress upon jurors that no one other than the defendant could be the source of the print. That is, once the examiner in our study offered his opinion that the defendant was the source of the print, it made no difference to our jurors whether that source claim was stated as a source opinion, a source opinion bolstered by a reference to “a reasonable degree of scientific certainty,” or some form of “individualization” conclusion. Presumably, then, people process these descriptions heuristically, and reason that the expert is simply telling them: “It’s the defendant’s fingerprint: period.” However, when the expert in our study offered a weaker, and more scientifically justifiable conclusion – one that left open the possibility that there are others besides the defendant who may be the source (see Conditions 1 and 2) – our jurors assigned less weight to the fingerprint evidence.

If the pattern of results we observed holds true across domains, then reform efforts that focus not on barring source conclusions or statements of identification but solely on eliminating the purely bolstering features of forensic match reports – features such as “to a reasonable degree of scientific certainty” or “to the exclusion of all other possible sources in the world” – may be ineffective. Although these claims may be unscientific and unhelpful, banning such language from the courtroom may have little practical effect on how jurors think about and use the forensic evidence they hear. If courts will not allow source attribution statements unless and until scientists can offer compelling scientific data that support such statements, then source attribution statements in any form should be prohibited at trial. In contrast, moving forensic experts toward more conservative, scientifically defensible claims such as “the defen-

dant cannot be excluded as a possible contributor of the print,” could represent an important change.³¹

Jurors in our study also assigned relatively less weight to fingerprint evidence when the Army’s language was used (“the likelihood that we would observe this degree of correspondence when two impressions are made by different sources is considered extremely low”).³² Here, as well, the perception of lower probative value probably reflects an understanding that the examiner’s statement does not preclude the reasonable possibility that people other than the defendant might have prints that matched the latents in the case.

The fact that cross-examination on the shortcomings of the forensic conclusion had no impact on our jurors is disheartening. But this result is not entirely surprising. Koehler’s results from a 2011 shoeprint study are similar.³³ He found that defense attorney cross-examination of a shoeprint expert had no effect on his mock jurors, even when that cross-examination revealed important risks that were ignored by the match statistic provided.³⁴ But it is important to remember that our cross-ex-

amination was only cursory and in print. It is possible that a well-tailored live cross-examination would be more effective.

Meaningful reform related to the way fingerprint evidence is reported should bring with it an acknowledgment that the available science does not enable examiners to prove that only one person could be the source of an unknown print. Source conclusions, including those that imply a kind of objective certitude (such as “individualization”) are little more than the subjective, untested opinions of examiners. In the words of the respected forensic scientist David Stoney, such conclusions represent “a leap of faith . . . a jump, an extrapolation, based on the observation of highly variable traits.”³⁵

Squaring scientific accuracy with public understanding of the value of forensic science evidence will require a greater focus on empirical research, both to explore further the scientific basis of fingerprint analysis and to identify ways to convey accurately what the science has to offer and its associated uncertainties. We see no place in this endeavor for individualizations and untested source opinions.

Joseph B.
Kadane &
Jonathan J.
Koehler

AUTHORS’ NOTE

This essay was presented at the *Dædalus* authors’ conference in Cambridge, Massachusetts, in July 2017, and we gratefully acknowledge the many helpful comments made by the conference participants. We also thank Shari Diamond, Brandon Garrett, and Rick Lempert for their helpful comments on previous drafts.

The material presented here is based upon work partially supported under Award No. 70NANB 15 H176 from the U.S. Department of Commerce, National Institute of Standards and Technology. Any opinions, findings, or recommendations expressed in this material do not necessarily reflect the views of the National Institute of Standards and Technology, nor the Center for Statistics and Applications in Forensic Evidence. This research was also supported, in part, by the Northwestern Pritzker School of Law Faculty Research Program.

ENDNOTES

¹ Francis Galton, *Finger Prints* (New York: MacMillan, 1892).

² Federal Bureau of Investigation, *The Science of Fingerprints: Classification and Uses* (Washington, D.C.: U.S. Government Printing Office, 1985), iv.

- ³ 60 Minutes, "Fingerprints," January 5, 2003.
- ⁴ FBI National Press Office, "Statement on Brandon Mayfield Case," May 24, 2004, <https://archives.fbi.gov/archives/news/pressrel/press-releases/statement-on-brandon-mayfield-case>; and Office of the Inspector General, Oversight and Review Division, *A Review of the FBI's Handling of the Brandon Mayfield Case* (Washington, D.C.: U.S. Department of Justice, 2006).
- ⁵ National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* (Washington, D.C.: National Academies Press, 2009).
- ⁶ Spencer S. Hsu, "FBI Admits Flaws in Hair Analysis Over Decades," *The Washington Post*, April 19, 2015; "The Justice Department and FBI have formally acknowledged that nearly every examiner in an elite FBI forensic unit gave flawed testimony in almost all trials in which they offered evidence against criminal defendants over more than a two-decade period before 2000."
- ⁷ Although latent print examiners generally will not call an identification in cases where the print quality is extremely poor, it is worth noting that when identifications are called on a low-quality print, they are commonly called at the same 100 percent confidence level that is used for identifications that involve very high-quality latent prints.
- ⁸ ACE-V stands for Analysis, Comparison, Evaluation and Verification. President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (Washington D.C.: Executive Office of the President, 2016), 9.
- ⁹ Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts, "Accuracy and Reliability of Forensic Latent Fingerprint Decisions," *Proceedings of the National Academy of Sciences* 108 (19) (2004): 7733 – 7738. Apparently, these claims of extreme accuracy have made their mark on the public. A recent study found that the median estimate mock jurors provided for the false positive error rate associated with fingerprint analysis is 1 in 5.5 million. See Jonathan J. Koehler, "Intuitive Error Rate Estimates for the Forensic Sciences," *Jurimetrics Journal* 57 (2017): 153 – 168.
- ¹⁰ National Institute of Standards and Technology and National Institute of Justice, *Latent Print Examination and Human Factors: Improving the Practice Through a Systems Approach* (Gaithersburg, Md.: National Institute of Standards and Technology, 2012), 13 – 18, http://www.nist.gov/manuscript-publication-search.cfm?pub_id=910745.
- ¹¹ Lewis Carroll, *Through the Looking Glass* (Minneapolis: Lerner Publishing Group, 2002), 247. See Dawn McQuiston-Surrett and Michael J. Saks, "Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact," *Hastings Law Journal* 59 (2008): 1159 – 1189. "Forensic expert witnesses cannot simply adopt a term, define for themselves what they wish it to mean, and expect judges and juries to understand what they mean by it"; *ibid.*, 1163.
- ¹² Jonathan J. Koehler, "When are People Persuaded by DNA Match Statistics?" *Law and Human Behavior* 25 (5) (2001): 493 – 513; Jonathan J. Koehler, "The Psychology of Numbers in the Courtroom: How to Make DNA Match Statistics Seem Impressive or Insufficient," *Southern California Law Review* 74 (2001): 1275 – 1306; Jonathan J. Koehler and Laura Macchi, "Thinking About Low-Probability Events: An Exemplar Cuing Theory," *Psychological Science* 15 (8) (2004): 540 – 546; Jonathan J. Koehler, "Linguistic Confusion in Court: Evidence From the Forensic Sciences," *Journal of Law and Policy* 21 (2) (2013): 515 – 539, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2227645; Dale A. Nance and Scott B. Morris, "Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random-Match Probability," *The Journal of Legal Studies* 34 (2) (2005): 395 – 444; Jason Schklar and Shari S. Diamond, "Juror Reactions to DNA Evidence: Errors and Expectancies," *Law and Human Behavior* 23 (1999): 159 – 184; and William C. Thompson and Erin J. Newman, "Lay Understanding of Forensic Statistics: Evaluation of Random Match Probabilities, Likelihood Ratios, and Verbal Equivalents," *Law & Human Behavior* 39 (4) (2015): 332 – 349.
- ¹³ McQuiston-Surrett and Saks, "Communicating Opinion Evidence in Forensic Identification Sciences" [see note 11]; and Dawn McQuiston-Surrett and Michael J. Saks, "The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear," *Law and Human Behavior* 33 (5) (2009): 436 – 453.

- ¹⁴ Brandon Garrett and Gregory Mitchell, “How Jurors Evaluate Fingerprint Evidence: The Relative Importance of Match Language, Method Information, and Error Acknowledgment,” *Journal of Empirical Legal Studies* 10 (3) (2013): 484 – 511, 497.
- ¹⁵ *Ibid.*, 489, 495.
- ¹⁶ *Ibid.*, 497.
- ¹⁷ See National Research Council, *Strengthening Forensic Science in the United States* [see note 5]. The recent decision by current Attorney General Jeffrey Sessions not to renew the National Commission on Forensic Science (NCFS) – a thirty-member advisory panel convened during the Obama administration “to enhance the practice and improve the reliability of forensic science” – introduces a large dose of uncertainty into the forensic science reform movement. See U.S. Department of Justice Archives, “National Commission on Forensic Science,” <https://www.justice.gov/ncfs> (accessed April 28, 2017); and Erin E. Murphy, “Sessions is Wrong to Take Science Out of Forensic Science,” *The New York Times*, April 11, 2017. The NCFS had been the leading force in the forensic reform movement over the past few years.
- ¹⁸ Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* (Washington, D.C.: U.S. Department of Commerce, National Institute of Standards and Technology, 2012), 129.
- ¹⁹ Defense Forensic Science Center, *The Use of the Term “Identification” in Latent Print Technical Reports* (U.S. Army Criminal Investigation Command, 2015). At the time of writing, the U.S. Army appears to be in the process of replacing this language with language provided by statistical interpretation software known as FRStat. See, for example, Heidi Eldridge, “The Shifting Landscape of Latent Print Testimony: An American Perspective,” *Journal of Forensic Science and Medicine* 3 (2) (2017): 72 – 81, 80. See also Arizona Forensic Science Academy, Forensic Science Lecture Series Fall 2017 Workshop Announcement, https://www.azag.gov/sites/default/files/sites/all/docs/azfsac/Announcement%20Fall%202017%20Workshop_LPC.pdf.
- ²⁰ See Federal Rules of Evidence, Rule 702, Testimony by Expert Witnesses; and David A. Stoney, “What Made Us Ever Think We Could Individualize Using Statistics?” *Journal of the Forensic Science Society* 31 (2) (1991): 197 – 199.
- ²¹ Michael J. Saks and Jonathan J. Koehler, “The Individualization Fallacy in Forensic Science Evidence,” *Vanderbilt Law Review* 61 (1) (2008): 199 – 219.
- ²² U.S. Department of Justice, “Proposed Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline” (Washington, D.C.: U.S. Department of Justice, 2016).
- ²³ Simon A. Cole, “Who Speaks for Science? A Response to the National Academy of Sciences Report on Forensic Science,” *Law, Probability and Risk* 9 (1) (2010): 25 – 46 [see “effect individualizations” statement by FBI Examiner Melissa Gische on page 36]; and Peter E. Peterson, Cherise B. Dreyfus, Melissa R. Gische, et al., “Latent Prints: A Perspective on the State of the Science,” *Forensic Science Communications* 11 (4) (2009), <https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct2009/review>.
- ²⁴ Jonathan J. Koehler and Michael J. Saks, “Individualization Claims in Forensic Science: Still Unwarranted,” *Brooklyn Law Review* 75 (4) (2010): 1187 – 1208.
- ²⁵ As Rick Lempert pointed out to us (personal communication, June 26, 2018), even if there were data that showed that no two people had the same fingerprint, one could not conclude that no two people could leave the same print because latent prints are commonly partial or smudged. And because latent prints are incomplete or smudged in various ways, it is not clear how we might go about estimating the frequency of those prints.
- ²⁶ U.S. Department of Justice, “Uniform Language for Testimony and Reports Initial Draft for Public Comment,” <https://www.justice.gov/archives/dag/forensic-science> (accessed June 15, 2017).
- ²⁷ President’s Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts*, 149 [see note 8].

- ²⁸ Ulery et al., “Accuracy and Reliability of Forensic Latent Fingerprint Decisions” [see note 9]; and Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts, “Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners,” *PLOS One* 7 (3) (2012): e32800.
- ²⁹ Koehler, “Linguistic Confusion in Court” [see note 12].
- ³⁰ McQuiston-Surrett and Saks, “Communicating Opinion Evidence in Forensic Identification Sciences” [see note 11]; McQuiston-Surrett and Saks, “The Testimony of Forensic Identification Science” [see note 13]; and Garrett and Mitchell, “How Jurors Evaluate Fingerprint Evidence” [see note 14].
- ³¹ We favor supplementing such conclusions with data, derived from rigorous empirical studies, that will help legal decision makers gauge the probative value of the reportedly matching items. Once such data are collected, it may well turn out that many fingerprint matches are as probative with respect to the source question as are DNA matches (holding aside the issue of the risk of human error that may differ across methods and personnel).
- ³² Defense Forensic Science Center, *The Use of the Term “Identification” in Latent Print Technical Reports* [see note 19].
- ³³ Jonathan J. Koehler, “If the Shoe Fits, They Might Acquit: The Value of Shoeprint Testimony,” *Journal of Empirical Legal Studies* 8 (2011): 21 – 48.
- ³⁴ *Ibid.*, 39.
- ³⁵ Stoney, “What Made Us Ever Think We Could Individualize Using Statistics?” 198 [see note 20].