3-18-2019

The impact of proficiency testing information and error aversions on the weight given to fingerprint evidence

Gregory Mitchell
University of Virginia

Brandon L. Garrett
Duke University

The impact of proficiency testing information and error aversions on the weight given to fingerprint evidence

Abstract

Fingerprint examiners regularly participate in tests designed to assess their proficiency. These tests provide information relevant to the weight of fingerprint evidence, but no prior research has directly examined how jurors react to proficiency testing information. Using a nationally representative sample of American adults, we examined the impact of proficiency testing information on the weight given to the opinions of fingerprint examiners by mock jurors considering a hypothetical criminal case. The fingerprint examiner's level of performance on a proficiency test (high, medium, low, or very low), but not the type of error committed on the test (false positive identifications, false negative identifications, or a mix of both types of error), affected the weight that jury-eligible adults gave to an examiner's opinion that latent fingerprints recovered from a crime scene matched the defendant's fingerprints, which in turn affected judgments about the defendant's guilt. Jurors who had no information about proficiency gave similar weight to the testimony as jurors exposed to highly proficient examiners, suggesting that jurors assume fingerprint examiners perform at high levels of proficiency unless informed otherwise. We also found that a plurality of Americans deems false acquittals just as aversive as false convictions and a significant minority deems false acquittals more serious. These differences in error aversions predicted differences in evidentiary assessments, suggesting that error aversions of jurors may play an important role in criminal trials.

Disciplines

Forensic Science and Technology

Comments

This is a manuscript of an article published as Mitchell, Gregory, and Brandon L. Garrett. "The impact of proficiency testing information and error aversions on the weight given to fingerprint evidence." *Behavioral sciences & the law* 37, no. 2 (2019): 195-210. Posted with permission of CSAFE.

THE IMPACT OF PROFICIENCY TESTING INFORMATION AND ERROR AVERSIONS ON THE WEIGHT GIVEN TO FINGERPRINT EVIDENCE

Cite as: Gregory Mitchell & Brandon L. Garrett, *The Impact of Proficiency Testing Information and Error Aversions on the Weight Given to Fingerprint Evidence*, 37 BEHAVIORAL SCIENCES & THE LAW 195-210 (2019).

Published version available at: https://onlinelibrary.wiley.com/doi/10.1002/bsl.2402

The Impact of Proficiency Testing Information and Error Aversions on the Weight Given to Fingerprint Evidence

Abstract

Fingerprint examiners regularly participate in tests designed to assess their proficiency. These tests provide information relevant to the weight of fingerprint evidence, but no prior research has directly examined how jurors react to proficiency testing information. Using a nationally representative sample of American adults, we examined the impact of proficiency testing information on the weight given to the opinions of fingerprint examiners by mock jurors considering a hypothetical criminal case. The fingerprint examiner's level of performance on a proficiency test (high, medium, low, or very low), but not the type of errors committed on the test (false positive identifications, false negative identifications, or a mix of both types of errors), affected the weight that jury-eligible adults gave to an examiner's opinion that latent fingerprints recovered from a crime scene matched the defendant's fingerprints, which in turn affected judgments about the defendant's guilt. Jurors who had no information about proficiency gave similar weight to the testimony as jurors exposed to highly proficient examiners, suggesting that jurors assume fingerprint examiners perform at high levels of proficiency unless informed otherwise. We also found that a plurality of Americans deems false acquittals just as aversive as false convictions and a significant minority deems false acquittals more serious. These differences in error aversions predicted differences in evidentiary assessments, suggesting that error aversions of jurors may play an important role in criminal trials.

THE IMPACT OF PROFICIENCY TESTING INFORMATION AND ERROR AVERSIONS ON THE WEIGHT GIVEN TO FINGERPRINT EVIDENCE

I. INTRODUCTION

Following *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) and subsequent revisions to Federal Rule of Evidence 702, judges must ensure that an expert relies on reliable methods and principles to reach conclusions. The error rate associated with an expert's method is an important consideration when determining whether the output of the expert's method is sufficiently reliable to be admitted into evidence (Koehler, 2008; Meixner & Diamond, 2014). If the expert's opinions are admitted, information about the error rate associated with an expert's method is relevant to determine the probative value of the expert's opinions. Jurors inevitably make assumptions about the fallibility of an expert's method, and, when empirical information is available about the true error rate, this information should be admitted to avoid incorrect assumptions and also to enable jurors to make comparisons across experts who utilize different methods with different error rates (as permitted in federal court by Federal Rule of Evidence 104(e)).

In many fields, experts take proficiency tests designed to measure their levels of accuracy and reliability as part of their regular training and accreditation.[1] For instance, many fingerprint examiners associated with state crime labs participate in annual proficiency testing. In these tests, the examiners are given a case file containing latent fingerprints impressions hypothetically taken from a crime scene along with rolled fingerprint impressions from persons who may have left the latent fingerprints. The examiner's task is to determine whether the latent prints provide sufficient information for identification purposes and, if so, whether the latent prints match any of the rolled prints. The test administrator then scores the test answers for number of latent prints deemed unusable, the number of hits (correct matches of latent prints to exemplar prints), correct rejections (correct conclusions of no match), misses (incorrect conclusions of no match), and false alarms (incorrect matches of latent prints to exemplar prints) (see Garrett & Mitchell, 2018, for a fuller discussion of proficiency testing in the context of fingerprint examinations). In practice, the proficiency tests currently commonly used are not blind, not demanding, and may not sufficiently assess the skill of an expert (Garrett & Mitchell, 2018). However, rigorous proficiency testing could provide expert-specific information about the reliability of an expert's work.

Evidentiary rules do not prescribe specific levels of proficiency that must be achieved to qualify a witness as an expert, nor have judges demanded rigorous proficiency information be provided in litigation. We argue that the level of proficiency demanded for

[1] In 2014, 98% of all publicly funded crime laboratories in the U.S. engaged in some form of proficiency testing of their analysts. *See* U.S. DEPARTMENT OF JUSTICE (2014). Most (95%) employed declared tests in which participants knew they were participating in a test. A smaller percentage used blind testing (10%), with participants not knowing their work was part of test, or random case reanalysis (35%), with prior samples being re-examined by new analysts to measure consistency in conclusions.

admissibility of an expert's opinion should depend, in part, on the degree to which judges believe that jurors will properly utilize error rate information when determining the weight to give to an expert's opinion. If jurors properly understand and utilize proficiency testing results when evaluating expert evidence, then judges should be more willing to admit expert opinions along with error rate information derived from proficiency testing. If jurors fail to understand and take into account error rate information, then greater care in admissibility determinations should follow, due to fear that jurors will over-value the expert's opinions relative to the risk of error.

The present research examines whether jury-eligible Americans understand and utilize proficiency testing results when determining how much weight to give to a forensic analyst's testimony. Specifically, we examine whether jury-eligible adults understand the significance of proficiency testing information for judging a fingerprint examiner's level of expertise and whether they take this proficiency information into account when determining how much weight to give to a fingerprint examiner's opinion that a latent print taken from a crime scene matches the defendant's fingerprints. We present mock jurors drawn from a representative sample of American adults with identification opinions of fingerprint examiners who vary in their levels of proficiency as part of the evidence in a hypothetical criminal case and ask these mock jurors to assess the strength of the evidence in the case and the expertise of the examiner. If jurors properly understand proficiency testing information, their judgments about the strength of the evidence and expertise of the examiner will increase as the examiner's proficiency increases. We also examine whether the types of errors committed on the proficiency test matter, asking whether the identification opinions of examiners who are more prone to false positives (i.e., erroneous identifications) are judged more cautiously than the opinions of examiners prone to false negatives (i.e., erroneous exclusions) or to a mix of both types of error.

In Part I, we situate our study in the context of prior research on the assumptions of jurors about the reliability and accuracy of forensic evidence and how jurors respond to information about forensic error rates. We describe our study's methods in Part II and present the study's results in Part III. In Part IV, we discuss the implications of our results for the use of expert testimony in criminal trials. While jurors have had some difficulty understanding other forms of information about forensic error rates, the participants in our study understood the significance of the proficiency information and deemed it important to their evidentiary assessments. Furthermore, our representative sample, when given no information about proficiency testing results, acted as if the expert was highly proficient, even though this assumption may be invalid in a number of cases. Accordingly, our study supports the view that proficiency testing information should be admissible at trial to ensure jurors have an accurate picture of the expert's fallibility and to help jurors determine what weight to give to the expert's opinions.[2]

[2] The results of proficiency tests taken by individual forensic analysts are not generally made public, but such test results often exist and can be obtained through pretrial discovery or questioning at trial (see Garrett & Mitchell, 2018). Alternatively, average field-wide proficiency levels can be estimated from publicly available data, but these estimates likely overestimate the true proficiency level of many analysts because of the relative ease of the proficiency tests compared to real case work (see Koehler, 2017). Our focus is on individual-level proficiency results because these results provide direct evidence of an analyst's level of expertise.

We also discuss an important secondary finding of our study: we asked participants whether they deemed false convictions or false acquittals a more serious error or whether they deemed the errors equally bad, providing data on the distribution of error aversions among the general public. Surprisingly high percentages of the public rate false acquittals to be worse than or as bad as false convictions, and the error aversions that jurors bring to trials appear to affect how they interpret and weigh evidence in a criminal trial. This finding has potentially broad implications for the criminal justice system and the practical meaning of the presumption of innocence.

II. THE PROBLEM OF ERROR IN FORENSICS

A. Juror Sensitivity to Forensic Error Rates

Although no prior study has examined how proficiency testing information for a particular expert affects jurors' assessments of that expert's opinions, prior research has examined how potential jurors respond to other sources of information about the error rates associated with a forensic identification method. In perhaps the earliest study of this kind, Koehler, Chia, and Lindsey (1995) presented jury-eligible participants with a summary of a murder case in which DNA evidence provided an important link between the defendant and the victim and found that information about random match probabilities, but not information about laboratory error rates, affected verdict choices. This finding was concerning because laboratory error rates tend to be larger than random match probabilities in the context of DNA evidence, yet potential jurors showed little sensitivity to laboratory error rates when assessing evidence of a supposed DNA match.

In contrast, Schklar and Diamond (1999) presented mock jurors with information about a hypothetical sexual assault crime, including expert testimony about (a) the probability that the defendant's DNA matched DNA recovered from the victim, (b) the probability of a random match, and (c) the probability of a laboratory error during the DNA matching process based on proficiency testing of the laboratory. They found that the risk of laboratory error reduced the likelihood participants would vote to convict the defendant, more so than the risk of a random match. They also found that the laboratory error rates disclosed by the expert were lower than the error rates expected by many of the participants. Thus, Schklar and Diamond (1999) found that laboratory error rate information may reduce *or bolster* confidence in forensic evidence, depending on the juror's pre-existing expectancies about those error rates.

In another study of DNA evidence, Thompson, Kaasa and Peterson (2013) examined how potential jurors confronted with a rape case and DNA evidence responded to information about the random match probability as well as information about the "false report probability," which mirrors laboratory-error conditions from earlier studies. The random match probability was held constant (at one in one trillion), but three variations on the false report probability were compared: the DNA analyst testified than an incorrect match due to accidental contamination of the DNA sample from the crime was impossible, occurs with approximately one in 10,000 samples, or occurs with approximately one in 100 samples. In a first study, chances-of-guilt estimates and rates of convictions both varied inversely with false report probability, but in a second study where the defendant had a strong alibi, the analyst's admission of that accidental contamination was possible had no effect on chance-of-guilt estimates or conviction rates. Secondary data analysis suggested that participants were sensitive to the error information in the second study but were already so skeptical of the prosecution's case that the information had no marginal impact on evidentiary assessments.

In prior research, we found that a fingerprint examiner's admission that fingerprint identifications were sometimes erroneous and that his own identification opinion in the case at hand might be in error significantly reduced the weight given to fingerprint evidence in a criminal trial (Garrett & Mitchell, 2013). This result held whether the fingerprint expert raised the possibility of error during direct testimony or conceded that possibility on cross-examination. Because the great majority of our participants had great confidence in the reliability of fingerprint identifications, merely acknowledging the possibility of error altered expectations about the evidence and reduced confidence in it. Koehler (2012) found a similar discounting effect when a forensic expert admitted during direct examination that his shoeprint-based identification could be the result of an error during the matching process, but cross-examination highlighting the risk of error had no additional impact.

Other research has examined the impact of disclosing the limits of a forensic method, without focusing specifically on error rates associated with the method. McQuiston-Surrett and Saks (2009) found that explaining the subjective and unscientific nature of identifications based on hair comparisons *increased* the weight given to an expert's probability estimate of a match, suggesting that this information helped jurors better understand the evidence without increasing skepticism of the method. Koehler and colleagues (2016) informed participants that the methods used to make bite-mark and fingerprint identifications had either undergone no scientific testing or a great deal of scientific testing; in a first study using an MTurk sample and a written case description, participants viewed both types of forensic identifications as stronger evidence when told the methods had been scientifically tested, but evidentiary assessments by a second sample recruited from a jury pool that watched a realistic trial simulation were not affected by whether the method had been scientifically validated.

Collectively, these studies show that jurors often are sensitive to information about the error rates associated with a forensic method, but this sensitivity depends on the nature of the error information and the assumptions jurors have regarding a particular forensic specialty. Jurors are likely to find it easy to discount general concerns about forensic evidence, including general error rate information, when confronted with an experienced expert with an impressive background, or when other evidence in the case corroborates the expert's conclusions. Moreover, many jurors may fail to appreciate the significance of information about a method's scientific status.

However, error information derived from proficiency testing of the testifying expert should be easy to understand and difficult to ignore. First, proficiency testing results for fingerprint examiners are straightforward: they reveal the number of correct identifications, incorrect identifications, correct exclusions, and incorrect exclusions. No advanced understanding of statistics or science is necessary to understand a proficiency test's results. Second, proficiency testing results are highly probative of the risk of error in the case at hand because they provide direct empirical evidence of the witness's level of expertise and care in utilizing a particular forensic method. All things being equal, experts with higher scores on a proficiency test (i.e., fewer missed items) should engender greater confidence, but in the context of a legal case, the testimony of an expert prone to false negative errors who testifies for the prosecution may be given greater weight than an expert who makes objectively fewer errors but is prone to false positive errors. Also, as Schklar and Diamond (1999) showed in the context of DNA evidence, the effect of error rate information derived from proficiency tests will likely depend on the expectancies about forensic evidence that jurors bring to the trial. Proficiency testing results that reveal relatively poor performance compared to other fingerprint examiners may nonetheless bolster confidence in an examiner's opinion if those results exceed jurors' expectations about usual error rates.

B. How Proficient Are Fingerprint Examiners?

Logically, proficiency testing results set bounds on the levels of accuracy and reliability that forensic analysts can achieve in their work. The information value of these bounds depends on the degree to which the proficiency tests simulate real work conditions. If items on a proficiency test are more difficult than items encountered in real casework, then arguably the proficiency test's results have no relevance in court for admissibility or weight determinations because a forensic analyst could perform poorly on the proficiency test but still perform well in the field. If items on a proficiency test are easier than items encountered in real case work, however, the proficiency test's results set strong upper bounds on an analyst's hit rates and correct rejection rates and weak lower bounds on miss rates and false alarm rates. Strong performance on an easy test provides little guarantee of good work in the field, but poor performance on an easy test should be cause for concern about an analyst's fieldwork.

The consensus among forensic evidence scholars is that the fingerprint comparisons found on proficiency tests commonly used by American crime labs are considerably less challenging than fingerprint comparisons found in real case work (e.g., Cole, 2005; Koehler, 2017; President's Council of Advisors on Science and Technology ("PCAST"), 2016). Indeed, the head of the Collaborative Testing Service ("CTS"), a leading provider of forensic proficiency tests, has conceded that "easy tests are favored by the forensic community" (PCAST, 2016, p. 57). The latent fingerprint impressions used in proficiency tests tend to be higher quality than those obtained from a typical crime scene, and typically the tests are *declared tests* (i.e., known to be tests rather than real cases) that are not monitored by an independent administrator and thus may sometimes involve collaboration among analysts to arrive at answers to the test.

Notwithstanding the relative ease of the CTS proficiency tests compared to real case work, results from these tests are informative because they likely set a floor on error rates within a forensic field. For the years 1995 to 2016, published results of CTS fingerprint proficiency tests reveal false positive error rates that vary annually from 0 to

23% and false negative error rates that vary annually from 0 to 43% (Garrett & Mitchell, 2018).[3] These results show that both false positive and false negative errors are common on even relatively easy fingerprint proficiency tests, and the typically higher false negative rates found on these tests suggest that examiners assume a conservative posture with respect to positive identifications, because a false identification is seen as a serious error within the field.

Proficiency tests utilizing samples similar to those encountered in the field likewise find evidence of both false positives and false negatives among fingerprint examiners, with the latter error rates typically far exceeding the former, providing further evidence of a conservative response bias on declared proficiency tests. Haber and Haber (2014) tabulated error rates across several fingerprint proficiency studies and found false positive error rates ranging from 0 to 3% and false negative error rates ranging from 1 to 13% when only incorrect exclusions were counted, but ranging from 9 to 55% when combining incorrect exclusions with incorrect inconclusives (i.e., when a matching print is declared unsuitable for comparison). For instance, in two proficiency studies conducted by Tangen, Thompson and McCarthy (2011), trained examiners made few false positive errors (although some individuals made multiple false positive errors) but many false negative errors. Thompson and colleagues (2013) observed that examiners on average failed to declare as matches over 27% of the prints that previously had been declared matches in the real cases from which the test prints were drawn.

C. Overview of the Present Research

The present study examines whether potential jurors do in fact treat proficiency testing results as probative of a fingerprint examiner's expertise and, accordingly, give greater weight to the identification opinions of highly proficient examiners. To test this hypothesis, we presented participants with a description of a case in which fingerprint examiners of varying proficiency opine that the latent fingerprints recovered from a weapon used in an armed robbery matches the fingerprints of the defendant. Proficiency information is conveyed through disclosure of the results of a recent proficiency test taken by the testifying examiner, with the examiner in one condition having received a perfect score on the proficiency test, one having received a score of only 66%, and other examiners receiving scores between these extremes (scores of 98%, 92%, or 86%). We examined this wide range of proficiency levels because some examiners do score perfectly on existing tests and because we presently do not know the lower bound on the proficiency of testifying examiners. We posited a score of 66% to be near or below the minimal level of proficiency of an examiner who would be offered to testify; scores closer to chance performance are likely to be grounds for exclusion.[4]

[3] CTS only publicly reports aggregate test results (i.e., the number of errors made across all print comparisons); thus, the reported rates reflect error rates within the field rather than for individual analysts or laboratories.

[4] Some of these levels may strike one as quite low. However, courts often have not required disclosure of individual proficiency test results (see Garrett & Mitchell, 2018), and the number of errors observed on many of the proficiency tests that are given annually to examiners proffer suggest some examiners who testify may have relatively low proficiency scores. In any event, the primary rationale for including lower proficiency score levels was to see how jurors reacted to this information and to compare the treatment of

We also manipulated information about the nature of the errors committed on the proficiency test to examine whether potential jurors treat false positive errors as more significant than false negative errors when weighing a positive identification opinion (the fingerprint expert in our mock case always opines that the latent prints match the defendant's prints). We therefore varied whether the examiner's errors on the proficiency test were all false positives, all false negatives, or a mix of both errors. The error type variable permits examination of whether potential jurors are particularly skeptical of fingerprint examiners testifying for the prosecution who are prone to false positive errors and/or particularly accepting of fingerprint examiners testifying for the prosecution who are prone to false negative errors.

Although the criminal justice system treats false convictions as more serious than false acquittals, which explains the heightened burden of proof in criminal trials, individual jurors may not share this normative view. We asked participants whether they deem false acquittals or false convictions to be a more serious mistake or whether they deem the two mistakes as equally serious, so that we could examine whether individual differences in error aversions affect the impact of the errors committed by fingerprint examiners.

Finally, to gauge default assumptions about the proficiency of fingerprint examiners, we included a control condition in which a fingerprint examiner gave an identification opinion but about whom no proficiency testing information was disclosed. By comparing evidentiary assessments in this control condition to those in the various proficiency testing conditions, we are able to gauge when an examiner's proficiency level falls above or below the assumed level of proficiency of fingerprint examiners. For instance, if jurors assume perfect proficiency on the part of fingerprint examiners, then evidentiary assessments in the control and perfect proficiency conditions should not differ.

III. METHOD[5]

A. Participants

Most mock jurors studies use convenience samples, but we commissioned Qualtrics to recruit a nationally representative sample with respect to gender, race/ethnicity, age, income, and geographic location in the United States to increase the generalizability of our results. A total of 1,450 adults received approximately $3 each for their participation in the study, which took less than 15 minutes (average time to completion was just over nine minutes). Seven hundred participants self-identified as male (48.3%) and 750 as female (51.7%). Approximately twelve percent of the participants were aged 18 to 24, 19% 25 to 34, 18% 35 to 44, 19% 45 to 54, 17% 55 to 64, 12% 65 to 74, 2% 75 to 84, and less than 1% 85 or older. Nine hundred and eighty-seven participants self-identified as White (68.1%), 182 as Black or African-Americans (12.6%), 168 as Hispanic (11.6%), 55 as Asian (3.8%), 10 as American Indian or Alaska Native (7%), 5 as Pacific Islander or Native Hawaiian (.3%), and 43 as other (3%). Participants resided in the West (22.7%), Midwest

lower proficiency examiners to that of an examiner for whom proficiency level is unknown, to gauge assumptions about examiner proficiency.

[5] Study materials and data are publicly available at https://osf.io/r63tz/.

(21.5%), South (36.6%) and Northeast (19.2%) regions of the U.S. More participants reported consistently voting for Democratic (40.1%) than Republican (30.9%) candidates, with the remainder stating that they have no consistent preference between the two main political parties (28.8%). Five hundred and twenty-eight participants had served on a jury (36.4%), with most of the trials being criminal trials (78%).

B. Procedure and Materials

The experimental materials were created using Qualtrics' online survey service, and participants completed the experiment on their own computers or smart phones by accessing a link to the online survey. The survey software randomly assigned participants to one of 14 conditions: our two independent variables (five proficiency levels and three error types) produced 13 experimental conditions rather than 15 conditions because the condition in which the examiner received a perfect score on the proficiency test involved no errors (hence, the error type variable was not manipulated at this perfect proficiency level); the fourteenth condition was the control condition, in which participants received fingerprint evidence without proficiency testing information for the fingerprint examiner.

The study began by eliciting participants' informed consent and obtaining information on participant demographics, political affiliations, prior jury service, and answers to the multiple-choice version of the Berlin Numeracy Test along with two subjective numeracy questions (how good are you with percentages, and how helpful do you find graphs and tables in news stories) (Cokely et al., 2012).[6] Participants then received a description of a pending criminal case in which the defendant had been linked to a crime by fingerprint evidence:

> A convenience store was robbed. The robber wore a mask and showed a gun but did not fire the gun. When running from the store, the robber's hand caught on the door, causing him to drop the gun. No other person handled the gun before it was secured by the police. 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.
>
> A fingerprint examiner compared fingerprints taken from the handle of the gun dropped at the crime scene to the fingerprints taken from the Defendant on an inked card. The fingerprint examiner issued a report concluding that: **"A fingerprint recovered from the gun matched the right thumb of the defendant."**

[6] Only 21 participants (1.4%) selected correct answers for all four items on the Berlin Numeracy Test, and only 93 (6.4%) correctly answered three of the items; 375 (25.9%) correctly answered two items, 628 (43.3%) correctly answered one item, and 333 (23%) failed to answer any items correctly. In contrast, over 60% of the participants rated themselves as somewhat good to extremely good with percentages, and 90% of the participants stated that they find tables and graphs in newspapers to be somewhat or extremely helpful. Thus, while much of our sample subjectively rated themselves as numerate, our objective measure found much lower rates of numeracy.

As is common within mock juror studies (e.g., McQuiston-Surrett & Saks, 2009; Schklar & Diamond, 1999; Scurich & John, 2013; Thompson et al., 2013), we intentionally kept the case description simple and focused on the fingerprint evidence to avoid possible bolstering of the fingerprint evidence through other corroborating evidence (corroborating evidence can lend credibility to the fingerprint identification and allay any concerns potential jurors may have about the reliability of the fingerprint evidence). Furthermore, although fingerprint examiners employ a variety of terms to indicate source identification (see Table 1 in Garrett & Mitchell, 2013), we chose one of the simpler formulations, using "match" terminology to avoid the need for an extended discussion of the expert's testimony and because the different formulations tend to produce the same results (Garrett & Mitchell, 2013; see also Kadane & Koehler, 2018).

Participants in the control condition received only the above case description, but participants in the proficiency testing conditions received information about the examiner's performance on a recent test. For instance, in the high proficiency/mixed error condition, participants received the following additional information immediately after the above case description:

> In evaluating this evidence, it may be helpful to have this additional information about the fingerprint examiner:
>
> A proficiency test is a test where a fingerprint examiner is given 100 pairs of fingerprints to determine whether or not each of the two prints originated from the same source. The goal of this test is to assess how accurate and reliable a particular fingerprint examiner is at making fingerprint-based identifications. The administrator who prepares the proficiency test knows whether each pair of prints come from the same source or from different sources. Therefore, the test administrator can determine the percentage of test items in which the expert correctly concluded that the fingerprints did or did not match. It is common practice in many crime labs to have fingerprint examiners participate in proficiency tests of this kind.
>
> From sworn testimony in another case, we have information about results on a recent proficiency test for the fingerprint examiner who is involved in this case.
>
> In particular, this fingerprint examiner received a score of 98 percent on the proficiency test. This means that in 98 out of 100 trials on the proficiency test, this fingerprint examiner correctly concluded that the test prints did or did not match. For one pair of fingerprints, the examiner incorrectly concluded that the fingerprints matched when the fingerprints did not in fact come from the same source, and for one pair the examiner incorrectly concluded that the fingerprints did not match when the fingerprints did in fact come from the same source.

Here is a summary of this examiner's results on the proficiency test:

Accurate Identifications:	98 out of 100 pairs
Erroneous Identifications:	2 out of 100 pairs

Error Breakdown:

Mistaken Matches:	1
Mistaken Non-matches:	1

In the proficiency conditions, the examiner received a score of 100, 98, 92, 86, or 66 on the proficiency test, and in the conditions where errors were made, the errors were described as all mistaken matches, all mistaken non-matches, or an equal mix of mistaken matches and mistaken non-matches.[7]

After reading the case description and, as applicable, the proficiency testing information, participants completed the dependent measures: (a) a rating of the likelihood that the defendant left his fingerprints at the scene of the crime (answered on a scale ranging from 0% to 100% likely, in ten percent intervals, which means responses were on an eleven-point equal interval scale); (b) a rating of the reliability of the fingerprint evidence (answered on a six-point scale ranging from very reliable to very unreliable); (c) a rating of the level of expertise of the fingerprint examiner involved in the case (answered on a seven-point scale ranging from far above average to far below average, with average in the middle); (d) a rating of the likelihood the person arrested for the crime was in fact the robber (answered on a seven-point scale ranging from definitely is the robber to definitely is not the robber); (e) a rating of the strength of the prosecution's case (answered on a four-point scale ranging from extremely weak to extremely strong).[8]

Finally, participants indicated whether they believe each person's fingerprints and DNA are unique (both answered as yes, no, or probably), specified their view on the reliability of fingerprint evidence generally (answered on a six-point scale ranging from very reliable to very unreliable), and specified whether they believe that falsely convicting an innocent person or failing to convict a guilty person causes more harm to society or whether they believe the two errors cause equal harm.

IV. RESULTS

[7] One of the proficiency conditions used in the original study contained a typographical error in the paragraph describing the errors made by the examiner, but the information provided in the summary table was correct. We corrected this typographical error and submitted this corrected description to a sample of American adults recruited through Amazon's MTurk service. The MTurk sample and Qualtrics sample, the latter of which was exposed to the typo, did not differ in their mean ratings of the examiner's level of expertise. Accordingly, we report in the text the results for all of the proficiency conditions collected using the representative sample collected throught Qualtrics.

[8] An attention check item (in which participants were directed by the question to select a specific multiple choice answer) was included among the dependent measures; those who failed to answer correctly were excluded from the study.

Table 1 reports mean scores on all of our dependent measures by experimental condition. We first examined whether participants differed in their ratings of the expertise of the fingerprint examiner who performed differently on the proficiency test.[9] Expertise ratings did differ by proficiency level (β = -.42, p < .001), but expertise ratings did not differ by proficiency error type (β < .01) nor did the error type variable interact with the proficiency level variable (β < .01).[10] Examiners receiving a perfect score on the proficiency test were rated as significantly more expert than examiners scoring 98 on the test (t(435)=3.12, p = .002), who were rated as significantly more expert than those scoring 92 (t(624)=3.67, p < .001), who were rated as significantly more expert than those scoring 86 (t(592)=3.38, p = .001), who were rated as significantly more expert than those scoring 66 (t(606)=5.46, p < .001). Therefore, participants treated proficiency level—but not error type—as having informational value with respect to the examiner's level of expertise across all levels of proficiency.

[9] For analyses that examine the potential effects of the error type variable, we exclude the perfect proficiency and control conditions, which did not manipulate this variable.

[10] All statistical tests were conducted using IBM's SPSS Statistics program, Version 25. For most statistical tests, we present unstandardized coefficients from multiple regression analyses as the effect size measure; these coefficients indicate the amount by which a dependent variable changes if we change an independent variable by one unit keeping other independent variables constant. For some tests, we present correlation statistics as the measure of the strength of asssoction between two variables. We report results of linear regressions due to their ease of interpretation, but a number of our dependent measures were ordinal rather than continuous. Therefore we conducted the same analyses using ordinal regressions and observed the same pattern of results with one minor exception: one of the interaction contrasts was significant for one of the dependent measures in the ordinal but not the linear regression. We discount this finding because it was the product of multiple comparisons and no other interactions were observed in ordinal or linear regressions. As we discuss in the text, the error type variable deserves further study, but the present results suggest that proficiency level rather than error type matters most.

Table 1: Dependent variable means by condition

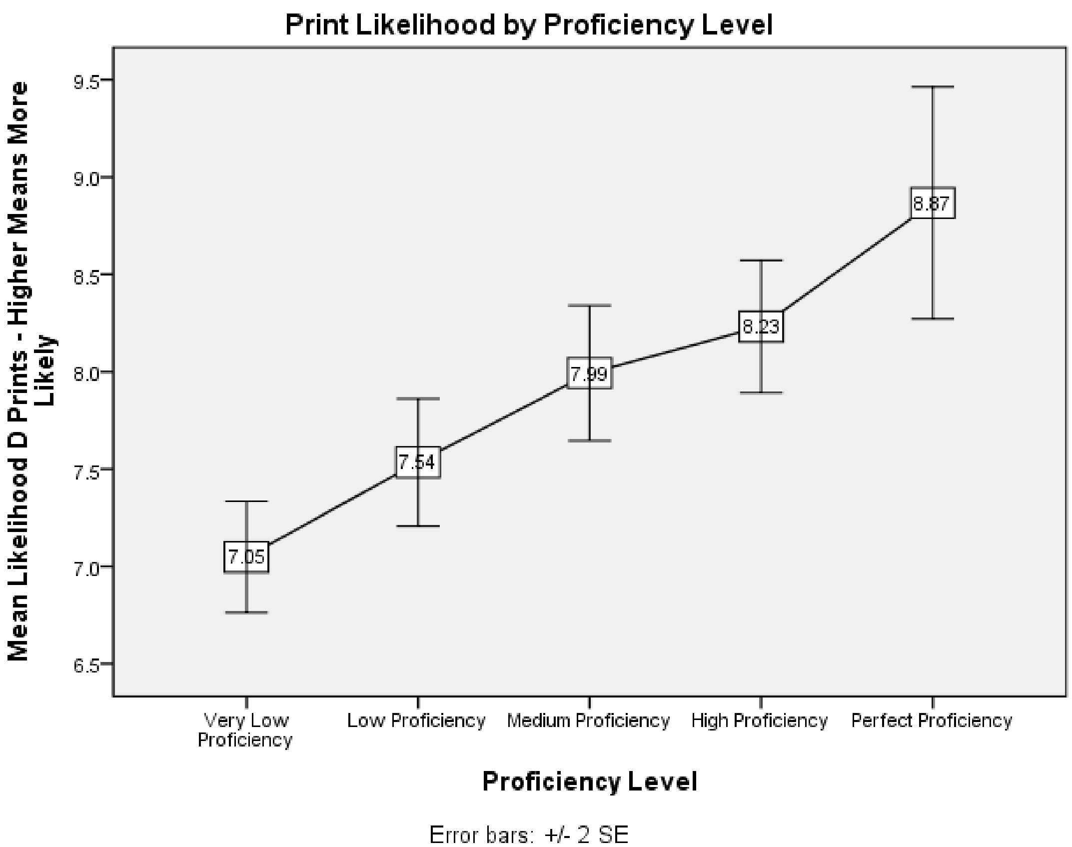
Condition	**Likelihood Defendant Left Fingerprints at Scene (1-11 scale, less to more likely)** ***M*** **(*n*, *SD*)**	**Reliability of the Fingerprint Evidence (1-6 scale, more to less reliable)** ***M*** ***(n, SD)***	**Expertise of Fingerprint Examiner (1-7 scale, far above to far below average)** ***M*** **(n, *SD*)**	**Likelihood Defendant Robber (1-7 scale, less to more likely)** ***M*** **(n, *SD*)**	**Strength of Prosecution Case (1-4 scale, weak to strong)** ***M*** ***(n, SD)***
Control	8.36 (110, 3.05)	2.02 (110, 1.11)	2.71 (110, 1.19)	5.00 (110, 1.51)	2.82 (110, .68)
100% Proficiency	8.87 (105, 3.05)	1.82 (105, .88)	1.70 (105, 1.06)	5.22 (105, 1.47)	3.07 (105, .75)
98% Proficiency					
2 False Positives (FP)	8.06 (112, 2.99)	2.03 (112, 1.01)	2.19 (112, 1.23)	5.05 (112, 1.57)	2.92 9112, .74)
2 False Negatives (FN)	8.21 (112, 3.21)	1.96 (112, .87)	2.10 (112, 1.04)	5.21 (112,1.33)	2.89 (112, .70)
1 FP, 1 FN	8.44 (108, 3.12)	1.97 (108, 1.01)	1.99 (108, 1.12)	5.13 (108, 1.55)	3.03 (108, .69)
92% Proficiency					
8 FP	7.95 (106, 3.17)	2.35 (106, 1.02)	2.45 (106, 1.27)	4.86 (106, 1.53)	2.76 (106, .72)
8 FN	7.88 (93, 2.98)	2.16 (93, 1.03)	2.49 (93, 1.33)	5.05 (93, 1.36)	2.86 (93, .67)
4 FP, 4 FN	8.15 (95, 2.76)	2.27 (95, 1.02)	2.38 (95, 1.14)	4.88 (95, 1.46)	2.80 (95, .63)
86% Proficiency					
14 FP	7.72 (101, 2.69)	2.37 (102, .89)	2.73 (102, 1.20)	4.66 (102, 1.35)	2.69 (102, .60)
14 FN	7.17 (95, 3.11)	2.34 (95, 1.04)	2.69 (95, 1.26)	4.81 (95, 1.53)	2.77 (95, .71)
7 FP, 7 FN	7.69 (103, 2.70)	2.56 (103, .99)	2.92 (103, 1.25)	4.76 (103, 1.35)	2.58 (103, .59)
66% Proficiency					
34 FP	6.46 (101, 2.44)	2.84 (101, 1.28)	3.50 (101, 1.57)	4.42 (101, 1.32)	2.47 (101, .67)
34 FN	7.70 (106, 2.76)	2.48 (106, 1.05)	3.32 (106, 1.50)	4.73 (106, 1.34)	2.65 (106, .70)
17 FP, 17 FN	6.96 (101, 2.15)	2.88 (101, 1.03)	3.38 (101, 1.48)	4.62 (101, 1.09)	2.48 (101, .59)

The proficiency level information also affected the weight given to the examiner's testimony in the case at hand. Participant perceptions of the reliability of the fingerprint evidence and of the likelihood that the defendant's prints matched the latent prints found on the gun recovered from the crime scene (as the fingerprint examiner testified) depended on the examiner's proficiency level (β = -.24, p < .001, and β = .37, p < .001, respectively) but not the type of errors committed by the examiner on the proficiency test.

As illustrated in Figure 1, below, we observed a linear trend in mean ratings of the likelihood the defendant's prints were on the gun, but not every contrast produced significant differences in ratings: likelihood ratings given by participants who read testimony from the examiner who scored 100 on the proficiency test were significantly higher than the ratings in all other proficiency level conditions except for the examiner who scored 98 on the test; likelihood ratings given when the examiner scored 98 on the test were significantly higher than those given when the examiners scored 86 and 66 on the test; likelihood ratings given when the examiner scored 66 on the test were significantly lower than ratings in every other condition (p < .05 for all t-tests for the relevant comparisons).

These differences in likelihood ratings were quite substantial: for instance, the most common likelihood rating for participants reading testimony from the very low proficiency examiner was 60%, whereas the most common likelihood rating for participants reading testimony from the perfect proficiency examiner was 100% (chosen by more than half the participants in this condition).

Figure 1: Print likelihood ratings by proficiency level



Participants' ratings of the likelihood the defendant left his prints on the gun predicted judgments about the defendants' likelihood of conviction and the strength of the prosecution's case ($\beta = .25$, $p < .001$, $r = .51$, and $\beta = .11$, $p < .001$, $r = .48$, respectively). We also examined whether the error type variable might have directly affected judgments about the likelihood of conviction and case strength, even though it did not do so indirectly by its effects on judgments of the likelihood the defendant left his prints on the gun. For example, perhaps the prosecution's willingness to use a fingerprint examiner prone to false positive errors suggests more generally an aggressive prosecutorial stance despite weak evidence. However, we found no direct effect of the error type variable on ratings of the likelihood of conviction or case strength. In sum, participants were quite sensitive to the fingerprint examiner's level of proficiency in their assessments of the evidence and the strength of the case for conviction, but the type of errors committed by the fingerprint examiner on the proficiency test had no significant effect on any of these assessments.[11]

We next examined how much weight participants in the control condition gave the fingerprint evidence compared to those in the various proficiency level conditions. Participants who received the fingerprint evidence without information about the examiner's performance on the proficiency test rated the likelihood that the defendant left his prints on the gun no differently than participants in the perfect proficiency, 98% proficiency, and 92% proficiency conditions, but participants in the control condition rated that likelihood to be significantly higher than participants in the 86% proficiency ($t(407) = 2.47$, $p = .014$) and 66% proficiency conditions ($t(435) = 4.45$, $p < .001$). These comparisons suggest that jurors assume a high, but not perfect, level of proficiency on the part of fingerprint examiners who testify at trial.

We included all of our individual difference measures in regression equations with the proficiency level and error type variables to predict ratings of the likelihood the defendant left his prints on the gun and likelihood of conviction.[12] The error type variable continued to have no significant predictive value, while the proficiency level variable continued to be the best predictor of the weight given to the fingerprint evidence of all of the variables in the equation ($\beta = .41$, $p < .001$). However, some groups did show differential responses to the fingerprint evidence.

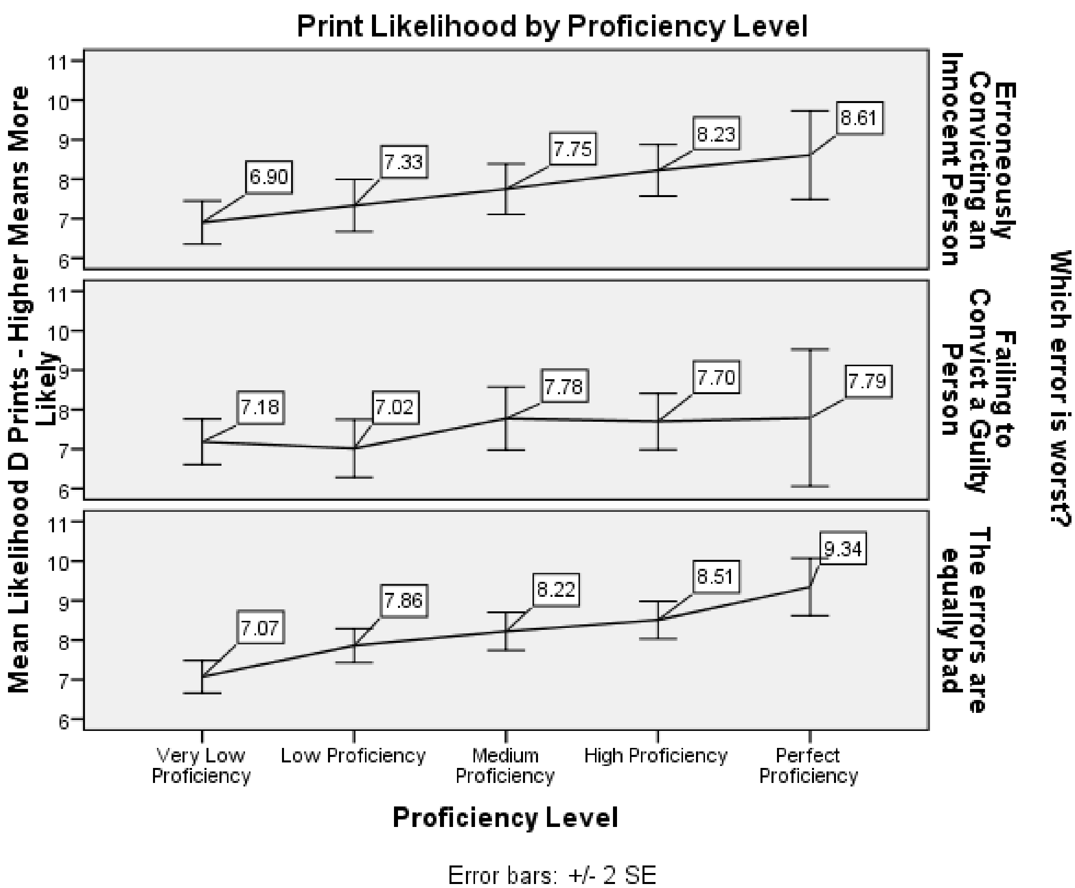
First, those with stronger aversions to false acquittals versus false convictions gave more weight to the fingerprint evidence regardless of experimental condition ($\beta = .20$, $p =$

[11] One might initially view our results as evidence of a demand effect at work (i.e., that mock jurors simply used the proficiency information because it was given to them). It is important to understand that such a characterization would not undercut the significance of the results because we are effectively testing whether jurors utilize information that they *should* utilize. Thus, our results are important whether one labels the result a demand effect or not. However, the full set of results casts doubt on the view that a simple demand effect was at work. First, the importance of the proficiency information depended on the level of proficiency: the impact of the proficiency information was not uniform across levels. Second, mock jurors did not distinguish among error types in using the proficiency information, which is inconsistent with the view that participants assumed that all information given to them should somehow alter their judgments. Third, we used a between-subjects design, lessening the likelihood that participants would infer what information the researchers deemed important or the hypothesized effects because they were not aware of changes in the variables across the experimental conditions.

[12] We excluded participants in the control condition from these analyses.

.039).[13] The proficiency level variable was least predictive of the ratings of likelihood that the defendant's prints were on the gun for those with stronger aversions to false acquittals (β = .22), compared to those who see false convictions and false acquittals as equally harmful (β = .50) and those with a stronger aversion to false convictions (β = .44). However, none of these slopes differed significantly, so we cannot say that the groups were differentially sensitive to the proficiency level information used in our experiment.[14] Figure 2 illustrates the relationship between proficiency level and likelihood ratings within each error aversion group. The flatter slope for persons most concerned about false acquittals suggests that this group may be more receptive to the prosecution's forensic evidence, an important possibility that deserves further study.

Figure 2: Print likelihood ratings by proficiency level and error aversion group



[13] We examined whether the distribution of error aversions varied by experimental condition (i.e., perhaps the experimental information affected responses to the error aversion question or perhaps one condition by chance had more participants with a particular error aversion), but the distribution of error aversions did not vary by experimental condition.

[14] In an additional test using only proficiency level and error aversions as predictors of judged likelihood that the defendant's prints were on the gun, both variables continued to produce main effects on judgments, but again the two predictor variables did not interact even when using this more focused comparison.

An interesting collateral finding is that many of our respondents saw false acquittals just as worthy of concern as false convictions. While many respondents rated false convictions to be the more harmful error (425, 29.3%), many more participants viewed false acquittals and false convictions to be equally harmful to society (709, or 48.9%), and a sizeable minority rated false acquittals as more harmful (315, 21.7%). Therefore, in our representative sample, the great majority did not share the criminal justice system's stronger aversion to false convictions but rather saw false acquittals as equally worthy of concern, and these concerns appear to have affected how participants interpreted the prosecution's evidence.

Women were more likely than men to rate false convictions and false acquittals as equally bad, while men were more likely to rate false convictions as worse ($\chi^2(2) = 33.30$, $p < .001$). Whites and minorities other than African Americans were more likely to rate the two errors as equally harmful, while a greater percentage of African American respondents rated false convictions as more harmful ($\chi^2(12) = 40.70$, $p < .001$). Those who had been arrested or who had a family member who had been arrested were more likely to rate false convictions as the more harmful error ($\chi^2(2) = 12.04$, $p = .002$).[15] Those who characterized themselves as very liberal were most likely to view false convictions as the more serious error ($\chi^2(8) = 25.35$, $p < .01$); in general, the more conservative the respondents, the greater the number rating the two errors as equally bad or rating false acquittals as the more serious error. As education increased, the percentage of participants rating false convictions as the more serious error increased ($\chi^2(12) = 27.65$, $p = .006$).[16]

Participants' level of objective numeracy ($\beta = .27$, $p = .003$) and race ($\beta = .27$, $p = .003$) also significantly predicted the weight assigned to the fingerprint evidence. Persons higher in objective numeracy gave the fingerprint evidence significantly more weight than those lower in objective numeracy. However, this relation held only at two levels of examiner proficiency (scores of 92% and 98%). In light of the small number of highly numerate participants in our sample (only 93 participants, or 6.4%, answered three of the numeracy items correctly and only 21, or 1.4%, answered all four items correctly) and the limited scope of this effect, further study is needed to determine whether highly numerate individuals interpret proficiency information differently than less numerate individuals. White and Asian participants on average gave higher ratings to the likelihood that the defendant's prints were on the gun than Blacks and Hispanics ($M = 8.01$ vs. 7.24, $t(1285) = 4.11$, $p < .001$), but all racial/ethnic groups were sensitive to the examiner's proficiency level when setting likelihood ratings (i.e., race/ethnicity and proficiency level did not interact). Note also that the effects associated with these individual difference variables were smaller than the effect observed for proficiency level. Therefore, these individual differences appeared to play less of a role in evaluations of the fingerprint evidence than did the examiner's proficiency level.

Finally, we examined general views on the reliability of fingerprint and DNA evidence. Consistent with prior research (Garrett & Mitchell, 2013), we found that the great

---

[15] African Americans and Hispanics reported higher rates of arrests than Whites and Asians (39% v. 23%; $\chi^2(1) = 41.61$, $p < .001$).
[16] Education level and political views were not correlated in our sample.

majority of respondents believed that each individual's fingerprints are unique (1,222, or 84.3%) or are probably unique (189, or 13%), while a few believed they that fingerprints are not individually unique (39, or 2.7%). A slightly smaller majority also believed that each individual's DNA is unique (1,188, or 81.9%) or is probably unique (173, or 11.9%), while a few believed an individual's DNA is not unique (89, or 6.1%).

Our participants viewed fingerprint evidence as generally reliable (M = 1.89, SD = .88, on a 1 to 6 scale ranging from very reliable to very unreliable), and believed that the average juror has considerable confidence in both fingerprint and DNA evidence (M = 2.89, SD = .68 and M = 3.23, SD = .73, respectively, with both rated on 1 to 4 scale ranging from no confidence in the evidence to a belief that the evidence is infallible). However, views on the general reliability of fingerprint evidence were correlated with certain group characteristics: men tended to rate fingerprint evidence as more reliable than women did (τ = .06, p = .023), older participants were more likely to rate the evidence's reliability higher (τ = -.05, p = .02), as were more conservative respondents (τ = -.07, p = .001) and those lower in subjective numeracy (τ = .10, p < .001). These differences across groups tended to be in the degree to which the evidence was rated as being reliable, with some more likely to choose very reliable over just reliable or somewhat reliable (e.g., M_{men} = 1.94 vs. M_{women} = 2.01, $t(1448)$ = -1.67, p = .095), and these effects were quite small across the various group lines. Thus, it appears that most jurors come to trial predisposed to trust fingerprint and DNA evidence and see such evidence as an excellent means of connecting particular individuals to crimes given the unique and generally reliable nature of such evidence. However, the intensity of these views may differ somewhat across groups.

V. IMPLICATIONS

Fingerprint examiners who score better on proficiency tests are less likely to make mistakes in their fieldwork than less proficient fingerprint examiners, holding other variables constant. This logic explains why accreditation bodies require proficiency testing and why scientists have recommended more rigorous proficiency testing (PCAST, 2016). Such testing is necessary to assess "the competence of particular examiners," but it is also an important way to engage in "validation and quality assurance" (Amercan Association for the Advancement of Sciences, 2017, p. 48). Judges should therefore take proficiency testing information into account when determining whether a forensic analyst should be admitted to testify as an expert witness. The ultimate factfinder should also take into account proficiency testing information when deciding how much weight to give to a forensic analyst's opinions at trial, giving greater weight to the opinions of more proficient experts.

The participants in our study consistently assigned greater weight to fingerprint match testimony coming from the more proficient fingerprint examiners. In making these weight assessments, participants consistently focused only on how many errors were made on the proficiency test and not on the type of errors made on the test. This result held true across the many diverse groups within our nationally representative sample, even for those persons most concerned about false convictions, who should be most sensitive to the prospect of a false identification.

These results suggest that jurors are quite sensitive to individualized information about a forensic analyst's error rate, and that this concern about the examiner's level of care and expertise does not vary with the nature of an expert's errors. Whether that result will hold for fields other than fingerprint comparisons, and whether there are some situations in which error type information will be independently influential, are questions that can only be answered with further studies. But the present study provides strong evidence that jurors can understand, and will use, proficiency testing information to assess an expert's testimony. Jurors recognize that not all experts are equally expert, and they condition their use of an expert's testimony on level of expertise.

The ability of our sample of potential jurors to properly understand proficiency testing information is shown by the way in which participant ratings of the fingerprint examiner's expertise tracked the examiner's level of proficiency. Participants understood that performance on a proficiency test is an empirical measure of expertise, and our participants used this proficiency information to calibrate their judgments about whether the defendant's fingerprints were on the gun used in the crime, and thus whether the defendant committed the crime. When jurors receive information about flaws or weaknesses in a forensic method or receive general information about a field's error rates, the juror cannot be sure how that information applies to the particular analyst in the case at hand. But when jurors receive information about the testifying expert's own performance on a proficiency test that simulates the task involved in the case at hand, the relevance of this information is easy to comprehend and hard to ignore.

The fact that jurors focused on proficiency level, and did not differentiate among types of errors on the proficiency tests, suggests that overall accuracy rather than avoidance of particular types of errors most concerned our participants. This finding suggests that CTS and other testing authorities should clearly report overall levels of accuracy for items for which an identification opinion is given and separately report the number of items for which specimens were rated inconclusive or not suitable for comparison. Given the conservative response bias observed on many proficiency tests that favor avoiding a match or no-match opinion (e.g., Haber & Haber, 2014), an important avenue of future study will be to examine how high rates of "deciding not to decide" (Dror & Langenburg, 2018) on proficiency tests affects the weight given to an examiner's opinions at trial. It may well be that jurors deem such behavior evidence of caution as opposed to evidence of test strategy or lack of confidence in the examiner's own ability. Future studies may therefore fruitfully examine the impact of levels of accuracy on attempted items as well as levels of opinion avoidance.

Our study also revealed that, absent specific information on a fingerprint examiner's level of proficiency, jurors are likely to assume that a fingerprint examiner has a high, if not perfect, level of proficiency. Evidentiary assessments in our control condition, in which participants received no proficiency testing information, mirrored evidentiary assessments in the conditions in which the examiner scored above 90% on a proficiency test. Presently, individual examiners' proficiency testing scores are not publicly available (and often courts will not order their discovery in litigation), but given the number of errors observed on many of the proficiency tests that are given annually to examiners, there are likely to be a number of fingerprint examiners testifying in trials who could not score above 90% on a realistic proficiency test.

Our results thus demonstrate not only the probative value of proficiency testing information in the eyes of potential jurors but also the need for individualized proficiency testing information to address possible misconceptions about a fingerprint examiner's error rate. In the past, some fingerprint examiners have claimed near infallibility in court (International Association for Identification, 2010; National Research Council, 2009). Individualized proficiency testing is needed to test the validity of such claims and to provide jurors with an empirically-derived measure of the examiner's actual level of care and expertise.

Although the effects of the proficiency level variable were robust across many different demographic and social groups, some groups did differ in their responses to the fingerprint evidence. Most notable was our finding that persons who believe that false acquittals do more harm to society than false convictions gave greater weight to the fingerprint examiner's testimony, but even these persons conditioned the weight assigned to the evidence on the examiner's level of proficiency. The number of people who have a stronger aversion to false acquittals is not negligible: a significant minority saw false acquittals as more harmful to society than false convictions, and almost half of our sample saw false acquittals and false convictions as equally harmful to society. These differences in error aversions are likely to have important effects on how evidence is interpreted and weighed in any criminal case and deserve further study. Jurors' error preferences may affect not only the weight given to evidence but also how the reasonable doubt and presumption of innocence instructions will be interpreted and applied.

VI. CONCLUSION

Jurors tend to assume that fingerprint examiners who testify at trials are highly proficient at making fingerprint identifications, yet the available proficiency testing data suggests that this assumption will not be true for all examiners (National Rsesarch Council, 2009; PCAST, 2016). This study demonstrates that jurors can properly understand proficiency information and use this information in logical ways *if that information is available to them*. Information from blind proficiency tests that mimic case conditions should be routinely collected across the forensic disciplines, and this information should be provided to judges and jurors so that potentially mistaken assumptions about proficiency can be replaced with individualized data on an expert's true level of proficiency.

VII. REFERENCES

American Association for the Advancement of Sciences (AAAS) (2017). Forensic Science Assessments: A Quality and Gap Analysis, Latent Fingerprint Examination.

ASCLD-LAB (2017). Proficiency Review Program Overview, http://www.ascld-lab.org/overview-proficiency-review-program/.

Cokely, E., Galesic, M., Schultz, E., Saimi, G., Garcia-Retamero, R. (2012). Measuring risk literacy: the Berlin numeracy test. *Judgment and Decisionmaking, 7,* 25-47.

Dror, I. E., & Langenburg, G. (2018). "Cannot decide": The fine line between appropriate inconclusive determinations versus unjustifiably deciding not to decide. *Journal of Forensic Sciences*.

Garrett, B. & Mitchell, G. (2013). How jurors evaluate fingerprint evidence: the relative importance of match language, method information and error acknowledgement. *Journal of Empirical Legal Studies*, *10*, 484.

Garrett, B. & Mitchell, G. (2018). The proficiency of experts. *University of Pennsylvania Law Review, 166*, 901-960.

Haber, R.N. & Haber, L. (2014). Experimental results of fingerprint comparison validity and reliability: A review and critical analysis. *Science & Justice, 54*, 375-389.

International Association for Identification. (July 16, 2010). IAI Resolution 2010-18.
Kadane, J.B., & Koehler, J.J. (2018). Certainty and uncertainty in reporting fingerprint evidence. *Daedalus, 147*, 119-134.

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

Koehler, J. J. (2017). Forensics or fauxrensics? Ascertaining accuracy in the forensic sciences. *Arizona State Law Journal*, *49*, 1369-1416.

Koehler, J., Schweitzer, N.J., Saks, M., & McQuiston, D. (2016). Science, technology or the expert witness: What influences judgments about forensic science testimony? *Psychology, Public Policy, and Law, 22,* 401-413.

Koehler, J.J. (2011). If the shoe fits they might acquit: The value of forensic science testimony. *Journal of Empirical Legal Studies, 8*, 21-48.

Koehler, J.J. (2008). Fingerprint error rates and proficiency tests: What they are and why they matter. *Hastings Law Journal*, *59*, 1077-1100.

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

McQuiston-Surrett, D., & Saks, M. (2009). The testimony of forensic identification science: what expert witnesses say and what factfinder hear. *Law and Human Behavior*, *33*, 436–53.

Meixner, J. & S.S. Diamond (2014). The hidden *Daubert* factor: How judges use error rates in assessing scientific evidence. *Wisconsin Law Review*, *2014*, 1063-1133.

National Research Council. (2009). *Strengthening forensic science in the United States: A path forward.* Washington D.C.: The National Academies Press.

President's Council of Advisors on Science and Technology. (2016). Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods, http://www.crime-scene-investigator.net/PDF/forensic-science-in-criminal-courts-ensuring-scientific-validity-of-feature-comparison-methods.pdf.

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

Scurich, N., & John, R. (2013). Mock jurors' use of error rates in dna database trawls. *Law and Human Behavior*, *37*, 424-431.

Tangen, J. M., Thompson, M. B., & McCarthy D. J. (2011). Identifying fingerprint expertise. *Psychological Science*, *22*, 995–997.

Thompson, W. C., Kaasa, S. O., & Peterson, T. (2013). Do jurors give appropriate weight to forensic identification evidence? *Journal of Empirical Legal Studies*, *10*, 359–397.

U.S. Department of Justice (2014), Census of publicly funded crime laboratories, https://www.bjs.gov/index.cfm?ty=dcdetail&iid=244.