Litigating Forensics Conclusions and Error Rates

Prof. Brandon Garrett

Feb. 2019

Duke University School of Law

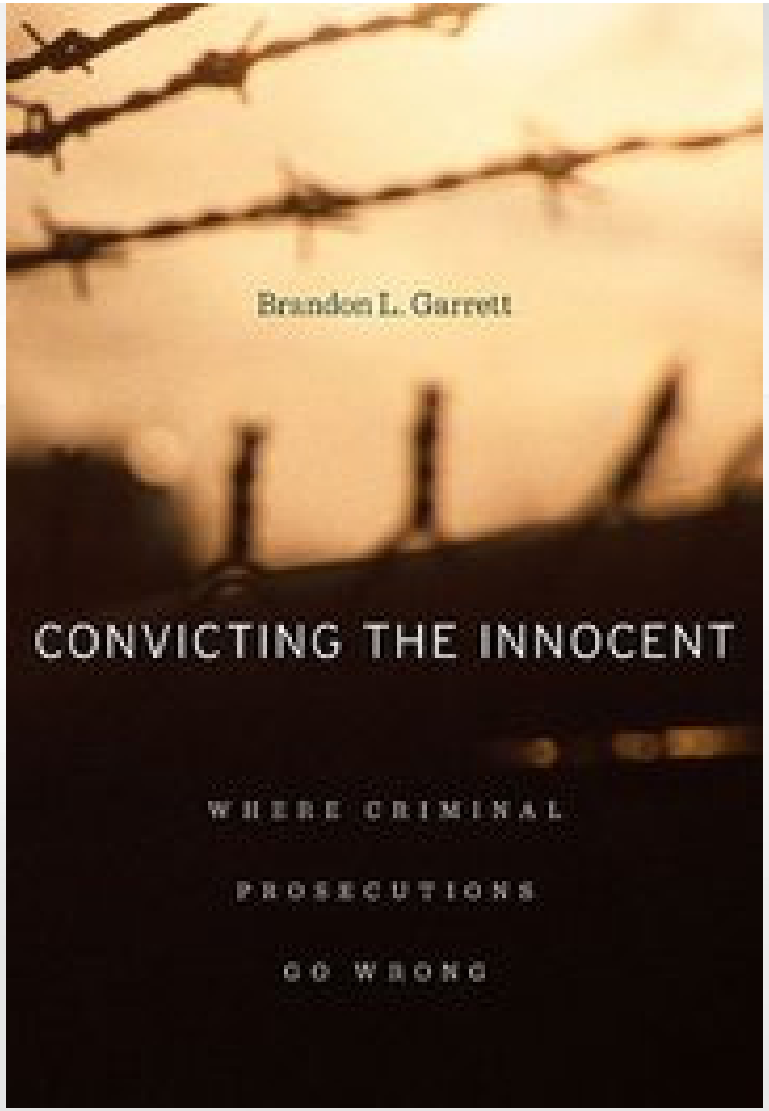
csafe
Center for Statistics and Applications in Forensic Evidence

1

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

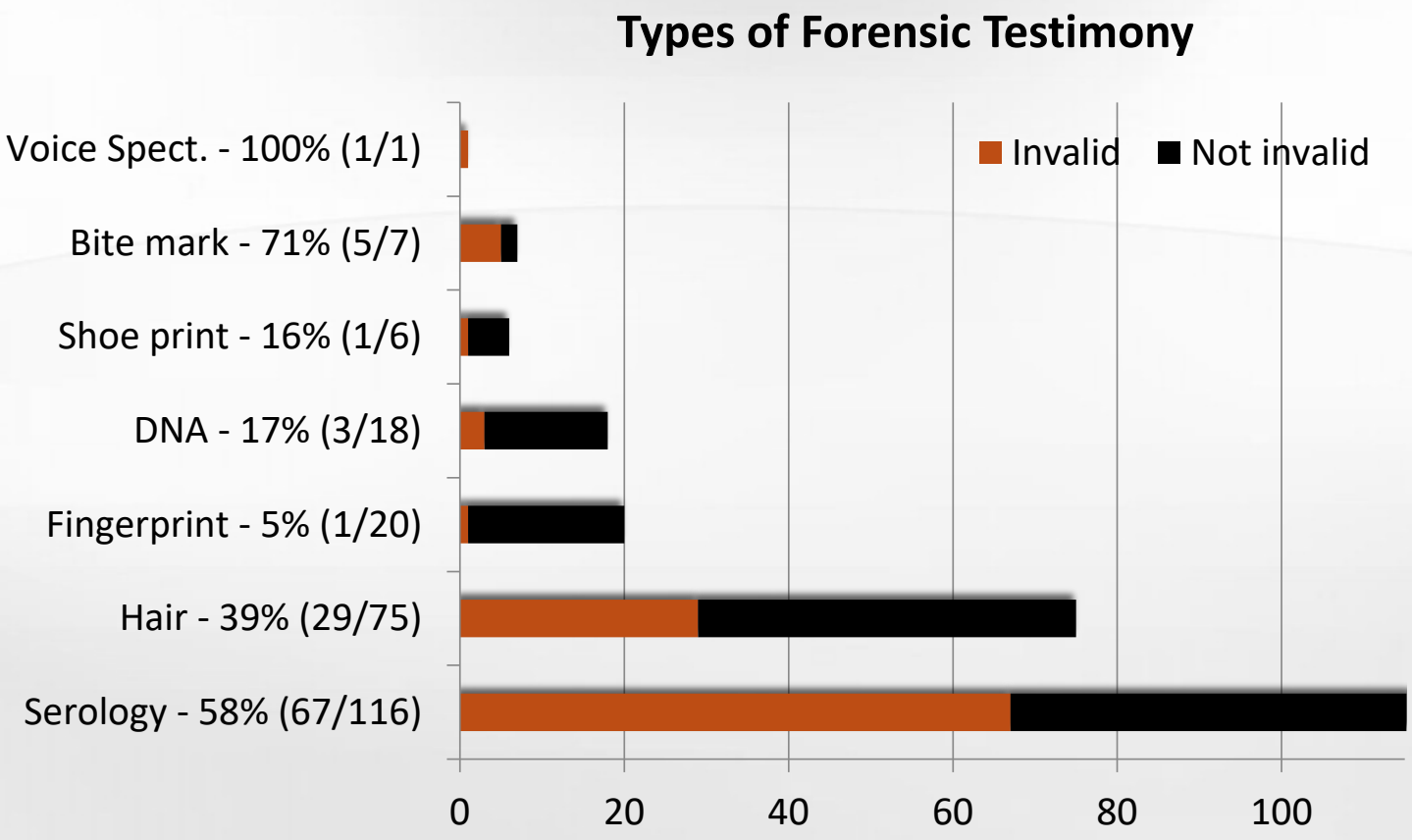
Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org



Data webpage
http://www.convictingtheinnocent.com

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org



Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

SWGFAST (3/13/2013)

- Individualization.
- Individualization is the **decision** by an examiner that there are **sufficient features in agreement** to conclude that two areas of friction ridge impressions **originated from the same source**. Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is **so remote that it is considered as a practical impossibility.**

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

International Association for Identification (IAI) (2009)

- "It is suggested that members not assert 100% infallibility (zero error rate) when addressing the reliability of fingerprint comparisons."
- "Members are advised to avoid stating their conclusions in absolute terms…"
  - Letter to all Members, Robert Garrett, IAI President, Feb. 19, 2009

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

PCAST Report (2016):

- Overall, it would be appropriate to inform jurors that **(1) only two properly designed studies of the accuracy of latent fingerprint analysis have been conducted and (2) these studies found false positive rates that could be as high as 1 in 306 in one study and 1 in 18 in the other study.**
- This would appropriately inform jurors that errors occur at detectable frequencies, allowing them to weigh the probative value of the evidence.
- "We also note it is conceivable that the false-positive rate in real casework could be higher than that observed in the experimental studies, due to exposure to potentially biasing information in the course of casework.
- And – "Proficiency testing is essential for assessing an examiner's capability and performance in making accurate judgments."

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

PCAST Report (2016):

- The PCAST report suggested forensic scientists use the term "**proposed identification**" in order to "appropriately convey the examiner's conclusion, along with the possibility that it might be wrong" (PCAST 2016, p. 45).

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

AAAS Report (2017) :

- Because there is no scientific basis for estimating the number of people who might be the source of a particular friction ridge print, we recommend that latent print examiners stop using the terms “identification” and “individualization.” These terms clearly imply that latent print examiners have the ability to single out the source of a print—to link it to a particular individual to the exclusion of any others.
- The term *identification*, proposed or not, implies an ability to limit the source of a friction ridge print to a single individual. That is an ability that latent print examiners cannot justifiably claim to have.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Defense Forensic Science Center (2015)

- The DFSC reporting statement (initially proposed by Swofford, 2015 and used until early 2017) is as follows:

- "The latent print on Exhibit ## and the record finger/palm prints bearing the name XXXX have corresponding ridge detail. The likelihood of observing this amount of correspondence when two impressions are made by different sources is considered extremely low" (Department of the Army, 2015).

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

DOJ Uniform Language

- 'Source identification' is an examiner's conclusion that two friction ridge skin impressions originated from the same source. This conclusion is an examiner's decision that the observed friction ridge skin features are in sufficient correspondence such that the examiner would not expect to see the same arrangement of features repeated in an impression that came from a different source and insufficient friction ridge skin features in disagreement to conclude that the impressions came from different sources.

- An examiner shall not assert that two friction ridge impressions originated from the same source to the exclusion of all other sources or use the terms 'individualize' or 'individualization.'
- An examiner shall not assert a 100% level of certainty in his/her conclusion, or otherwise assert that it is numerically calculated.
- An examiner shall not assert that latent print examination is infallible or has a zero error rate.
- An examiner shall not cite the number of latent print comparisons performed in his or her career as a measure for the accuracy of a conclusion offered in the instant case.
- An examiner shall not use the expressions 'reasonable degree of scientific certainty,' 'reasonable scientific certainty,' or similar assertions of reasonable certainty as a description ofthe confidence held in his or her conclusion in either reports or testimony unless required to do so by a judge or applicable law.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

NCSCL Language

- **11.1.7** The identifiable latent (fingerprint(s)/palmprint(s)/ impression(s)) was/were compared to Item (Item number) and was/were identified as having been made by the (finger of subject).
- An identification is defined as the decision by an examiner that there are sufficient features in agreement to conclude that two (2) areas of friction ridge impressions originated from the same source.
- Identification of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered a practical impossibility.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Stephan Cowans' Trial

Q And you look for points of identification that match up?

A Yes.

Q And how many total points of identification do you need to make a positive identification?

A Eight.

A That they were identical.

Q Whose print was it?

A Stephan Cowans'.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Testimony in State v. Wadrett

- "[T]he bullets from the victim's body and the casings from Carroll Avenue were fired from State's Exhibit 27, a firearm recovered in a field on Rick Boone Road."

- https://www.ncappellatecourts.org/show-file.php?document_id=229971

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

State v. McPhaul

- The examiner concluded that prints on the car and on the pizza and chicken boxes all were “identified” as coming from McPhaul.
- Going further still, the examiner stated that “[i]t was the left palm of Juan Foronte McPhaul that was found on the back fender portion of the vehicle.”
- And “[m]y conclusions, your Honor, is that the impressions made belonged to Mr. McPhaul.”
- The judge asked, “What did you do to analyze them?” and the examiner responded, “I did comparisons—side by side comparisons. ”
- She could not say what points were found on the prints, or what features were relied upon, what process were followed, or what the duration of the examination was.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

State v. Williams (COA 2018)

- "[T]he four evidence cartridge cases were fired from the same firearm as the test firings that I found. So they were fired from the same gun that I test fired."
- COA:
  - [W]hile Mr. Bishop did not qualify his opinion with "to a reasonable degree of certainty," he also never uttered the words "unique as to each gun that's made" or "exclusive identification" two phrases defendant refers to extensively in his brief as the alleged claims of certainty that amounted to false overstatements of reliability. In fact, it was defense counsel, not Mr. Bishop, who chose to use the exact phrases defendant challenges on appeal.
  - At no time, either on direct or cross-examination, did defense counsel object to any portion of Mr. Bishop's testimony or dispute the reliability of his expert opinion.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

But: Defense Experts

- Affirming exclusion of portions of defense expert testimony on self-defense reaction – "our review of the record in this case demonstrates that the trial court properly fulfilled its gatekeeping role."
- The trial court "carefully considered each aspect of the expert testimony that defendant wished to elicit"
- State v. McGrady, No. 72PA14 (NC 2016)

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Brandon L. Garrett & Chris Fabricant,

***The Myth of the Reliability Test*,**

86 Fordham L. Rev. 1559 (2018)

- *We assembled a collection of 229 state criminal cases that quote and in some minimal fashion discuss the reliability requirement.*
- *We find that in the unusual cases in which state courts discuss reliability under Rule 702 they invariably admit the evidence, largely by citing to precedent and qualifications of the expert or by acknowledging but not acting upon the reliability concern. In short, the supposed reliability test adopted in Rule 702 is rarely applied to assess reliability.*

Appendix II: State Rule 702 Adoption and Usage in Criminal Cases

State	Year Adopted	Cases Discussing Reliability
Alabama[165]	2012	1
Arizona	2012	17
Delaware	2001	4
Florida	2013	1
Georgia	2013	0
Indiana[166]	1994	18
Kansas	2014	1
Kentucky	2007	3
Louisiana	2014	5
Massachusetts[167]	n/a	6
Maryland	1993	1
Michigan	2004	59
Mississippi	2003	37
Missouri	2017	0
New Hampshire	2016	2
North Carolina	2011	14
Ohio[168]	1994	22
Oklahoma[169]	2013	3
South Dakota	2011	2
Utah[170]	2007	13
Vermont	2004	7
West Virginia	2014	2
Wisconsin	2011	11

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Experts May Invent Their Own Language

- The analyst concluded direct by stating that "Mr. Jackson is the person who bit this lady."
- The defense asked, "Is it your testimony… that based upon your analysis these bite marks in this case couldn't be made by anybody else?

  A. I never said that."
- There was no follow up – that concluded the cross.
- The State then elicited again that "there is no doubt in my mind that Willie Jackson is the individual who bit" the victim.

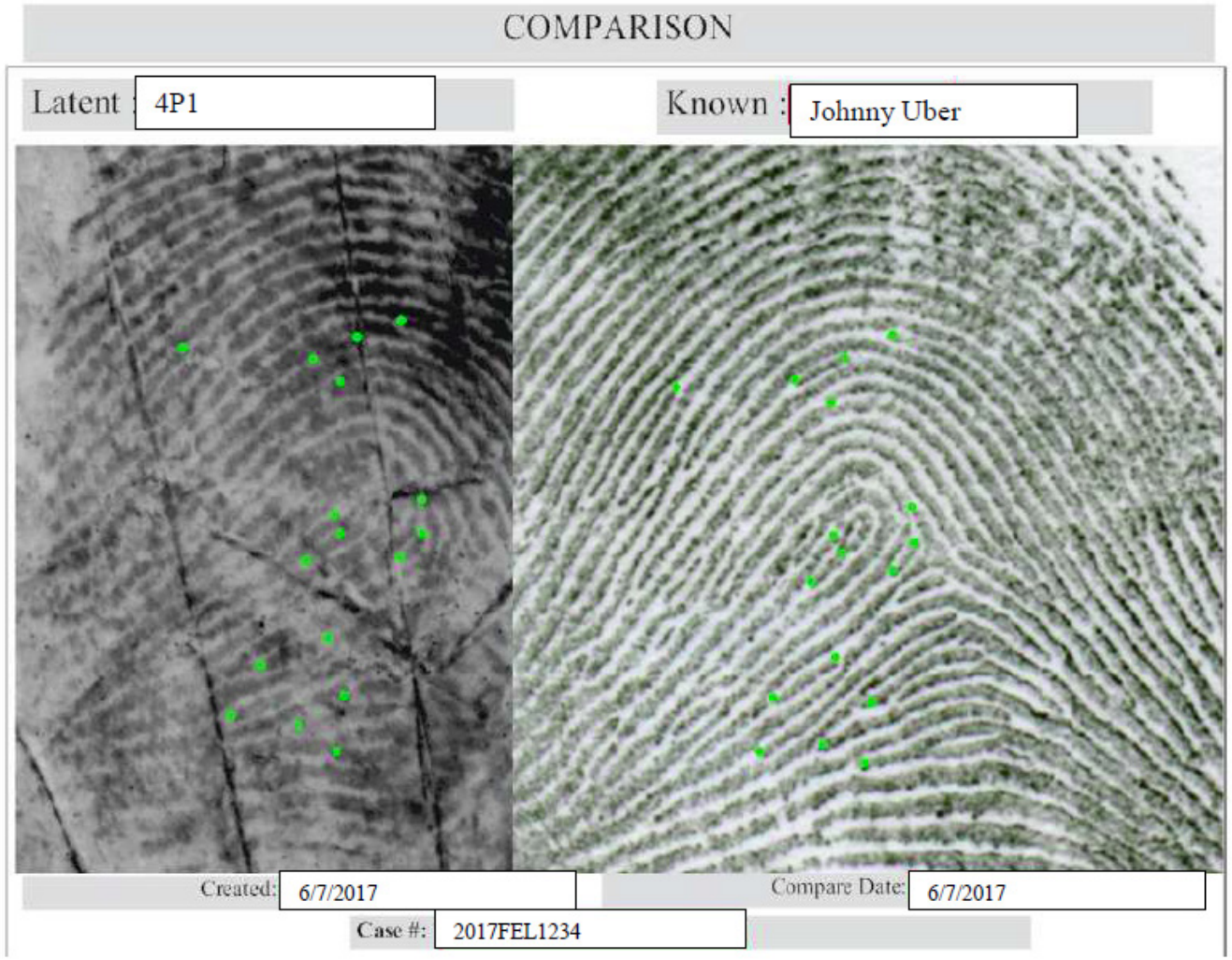
Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Cross that Elicits Added Exaggeration – the Trial of Glen Woodall

- "Q. Can you state objectively that that hair sample belonged to Glen Woodall?
- A. I would say…that the consistencies were 100 percent, and it is very highly likely that they came from the same individual."
- Q. But your answer is it was highly likely. You can't say it did, can you?
- A. There again, from the standpoint of scientifically stating from the characterization on the examination, I would say there was nothing to show me in the examination that they originated from another individual."

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Forensics Litigation Mock-Trial Casefile

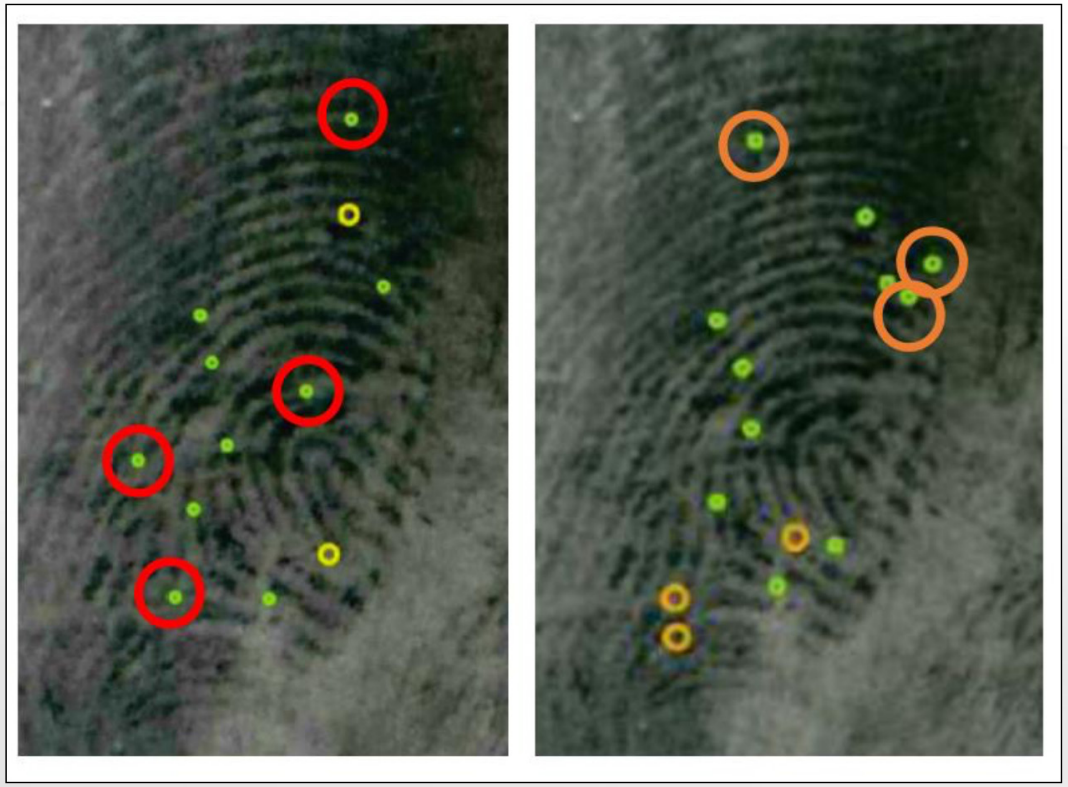


Verified: *MP*

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

UVA Law Student trial exhibit

A trial exhibit one of the teams prepared using case file images prepared and shared by the Virginia Department of Forensic Services.



Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

How are the prints marked?

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

How about Judges? Judicial Instructions?

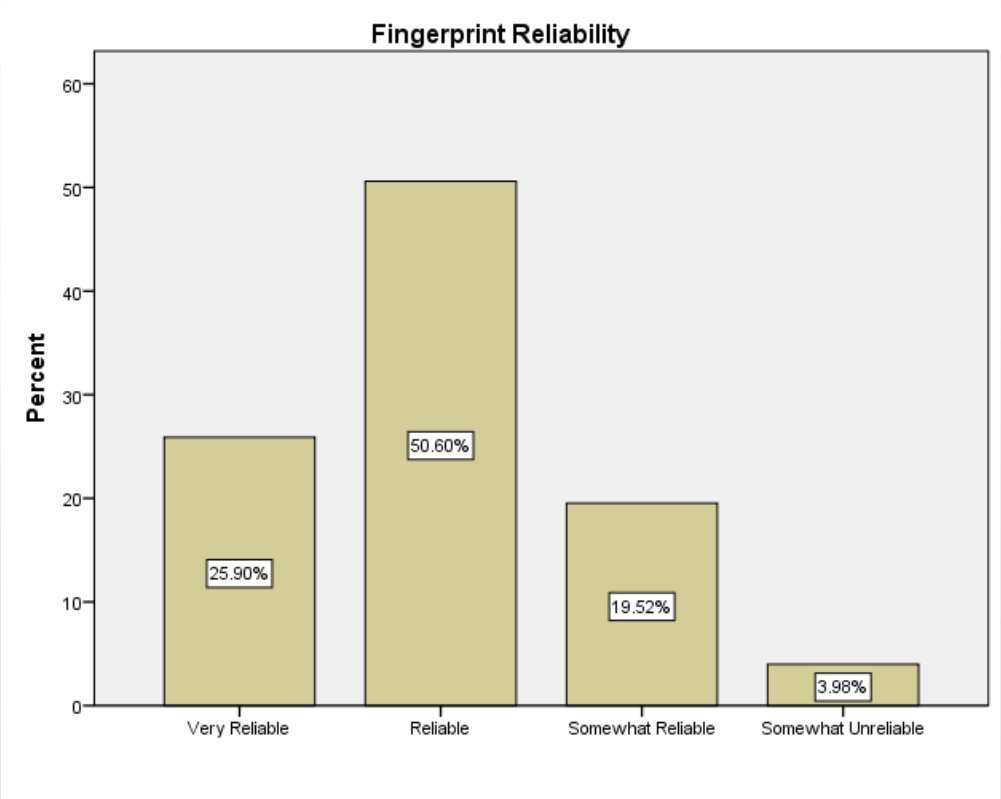
- Although "commissions come and go" there is "mounting judicial, and public concern…"

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

How do jurors evaluate such information?

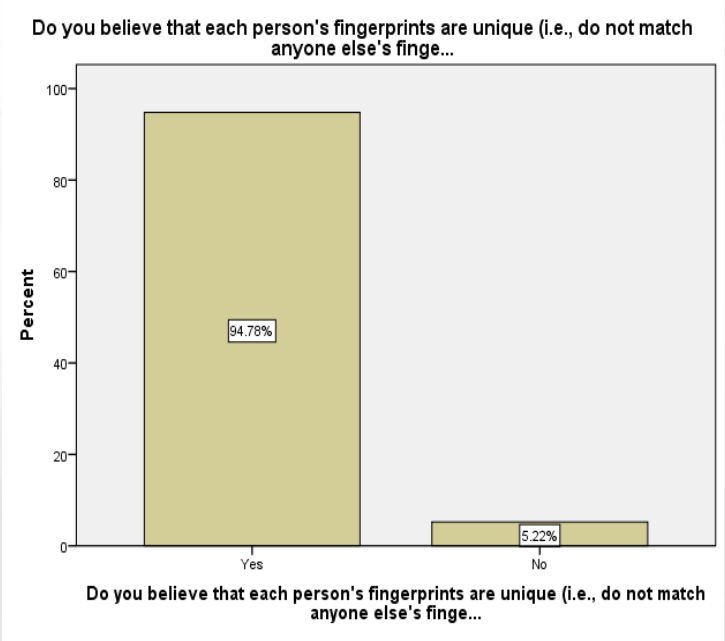
Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

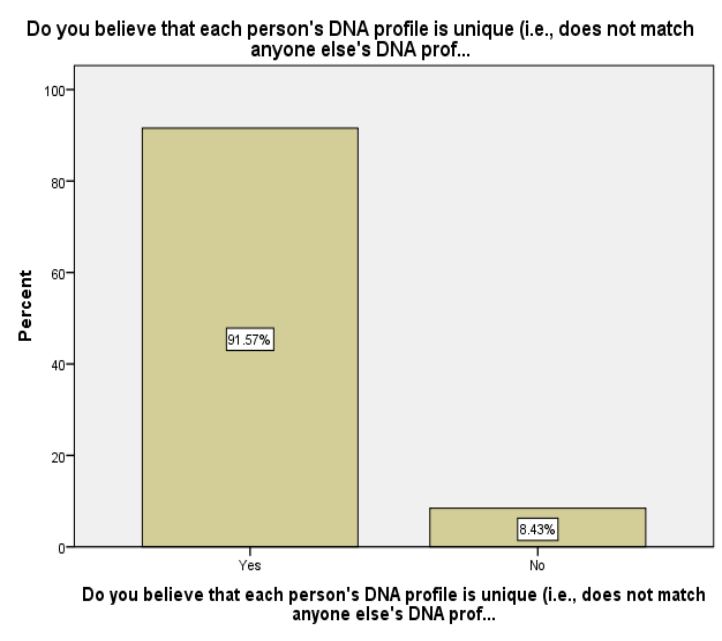
Lay Perceptions of General Fingerprint Reliability



Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Fingerprints vs. DNA





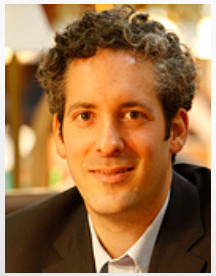
Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Prior work: *How Jurors Evaluate Fingerprint Evidence:*

The Relative Importance of Match Language, Method Information and ***Error Acknowledgement***

Brandon Garrett & Gregory Mitchell

10 JOURNAL OF EMPIRICAL LEGAL STUDIES 484 (2013)



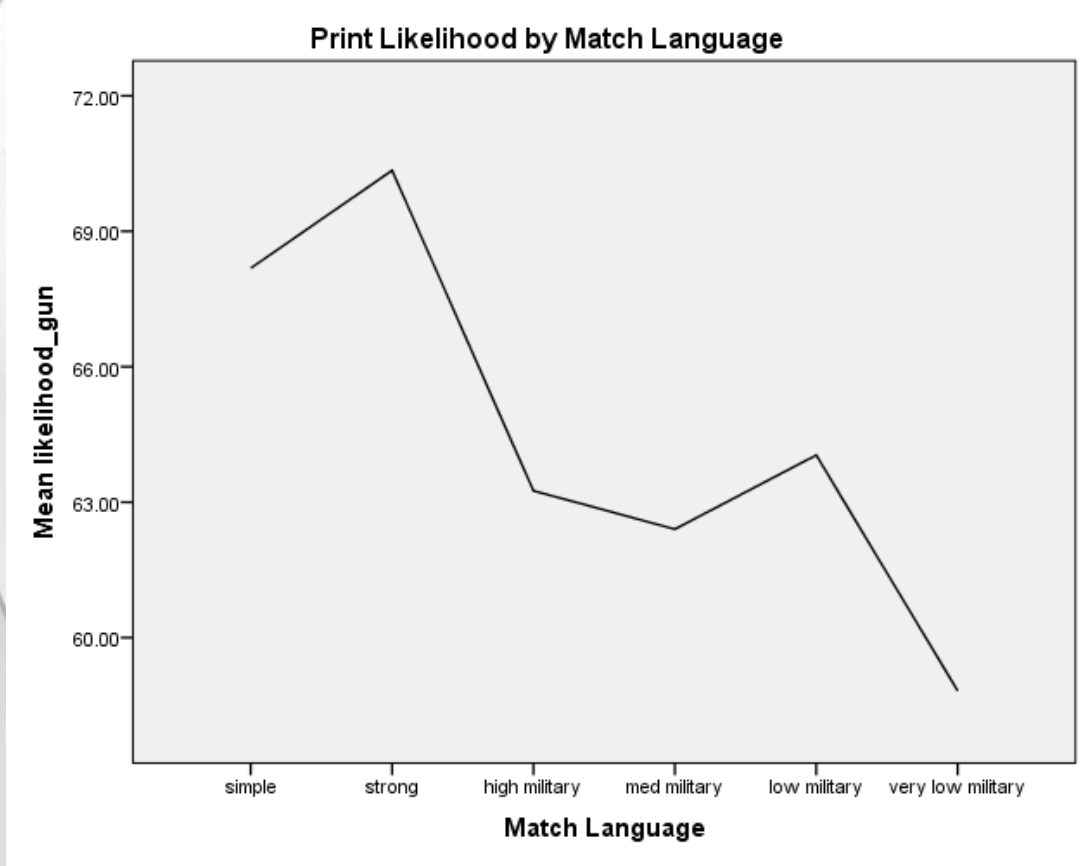
Fingerprint evidence benefits from **common beliefs and background assumptions** in uniqueness and reliability of fingerprint identification

Language may not be important once jury is told a match was made (Match = Match to Exclusion of All Others = 100% Certain = Other Source a Practical Impossibility….)

Error statements by forensic experts should be given greater attention by courts and researchers

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

New work: How do Jurors Evaluate FRSTAT Conclusions Concerning Fingerprint Evidence?



- The most probative FRSTAT conclusions were weighted as less strong than a traditional identification conclusion.
- The less probative FRSTAT conclusions were weighted less, but not much less, depending on their strength.
- Additional language describing the procedure did not impact results.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

PCAST Report on Verification

- It is important to note that, for a verification program to be **truly blind and thereby avoid cognitive bias**, **examiners cannot only verify individualizations**. As the authors of the FBI black-box study propose, "this can be ensured by performing verifications on a mix of conclusion types, not merely individualizations"—that is, a mix that ensures that verifiers cannot make inferences about the conclusions being verified.
- **We are not aware of any blind verification programs that currently follow this practice.**

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

(2) Introduce Error Rate Data: False positives

REPORT TO THE PRESIDENT

Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods

Executive Office of the President
President's Council of Advisors on
Science and Technology

September 2016

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

FBI Study: Ulery et al

- The study reported 6 false positive identifications among 3628 nonmated pairs that examiners judged to have “value for identification.” The false positive rate was thus 0.17 percent (upper 95 percent confidence bound of 0.33 percent). The estimated rate corresponds to 1 error in 604 cases, with the upper bound indicating that the rate could be as high as 1 error in 306 cases.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

More studies

- ***Tangen et al. (2011)***
- For the "similar" non-mated pairs, the experts made 3 errors among 444 comparisons; the false positive rate was 0.68 percent (upper 95 percent confidence bound of 1.7 percent), with the upper bound corresponding to 1 error in 58 cases.
- ***Miami-Dade study (Pacheco et al. (2014))***
- The false positive rate was 4.2 percent (upper 95 percent confidence bound of 5.4 percent). The estimated rate corresponds to 1 error in 24 cases, with the upper bound indicating that the rate could be as high as 1 error in 18 cases.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

PCAST Bottom Line:

- Overall, it would be appropriate to inform jurors that **(1) only two properly designed studies of the accuracy of latent fingerprint analysis have been conducted and (2) these studies found false positive rates that could be as high as 1 in 306 in one study and 1 in 18 in the other study.**
- This would appropriately inform jurors that errors occur at detectable frequencies, allowing them to weigh the probative value of the evidence.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

(3) Litigate Proficiency

- Seek discovery on proficiency
- Introduce proficiency data at trial

Year (Test #)	N of Prints	N of Test Takers	False Positive Rate	False Negative Rate (N)	Inconclusive Rate
1995 (9508)	7	156	22% (34)	43% (67)	3% (6)
1996 (9608)	11	184	8% (14)	N/A	N/A
1997 (9708)	11	204	10% (21)	28% (58)	N/A
1998 (9808)	11	219	6% (14)	N/A	35% (77)
1999 (99-516)	12	231	6% (14)	N/A	32% (75)
2000 (00-516)	10	278	4% (11)	N/A	N/A
2001 (01-516)	11	296	3% (8)	N/A	N/A
2001 (01-517)	11	120	20% (24)	N/A	N/A
2002 (02-516)	11	303	4% (13)	N/A	1% (2)
2002 (02-517)	10	146	3% (5)	4% (6)	1% (2)
2002 (02-518)	12	31	0% (0)	3% (1)	3% (1)
2003 (03-516)	10	336	1% (4)	8% (26)	N/A
2003 (03-517)	12	188	1% (5)	12% (22)	N/A
2003 (03-518)	9	28	7% (2)	11% (3)	N/A
2004 (04-516)	12	206	4% (12)	3% (7)	N/A
2004 (04-517-518)	15	259	6% (15)	2% (5)	N/A
2005 (05-516)	16	327	1% (3)	9% (28)	At least 1
2005 (05-517-518)	16	250	5% (12)	2% (6)	N/A
2006 (06-516)	15	333	23% (78)	3% (11)	N/A
2007 (07-516)	15	351	4% (14)	5% (18)	N/A
2007 (07-517-518)	15	315	4% (13)	14% (45)	6% (20)
2008 (08-516)	15	300	1% (3)	5% (14)	N/A
2008 (08-517-518)	15	391	1% (5)	2% (6)	1% (2)
2009 (09-516)	16	321	11% (35)	N/A	N/A
2009 (09-517-518)	16	419	1% (5)	2% (8)	1% (4)
2010 (10-516)	16	331	8% (26)	2% (5)	N/A
2010 (10-517-518)	16	463	13% (60)	N/A	N/A
2011 (11-516)	15	335	9% (30)	N/A	1% (3)
2011 (11-517-518)	16	478	4% (17)	0	.2% (1)
2012 (12-515-516)	16	350	3% (9)	2% (6)	N/A
2012 (12-517-518)	12	555	3% (16)	1% (8)	N/A
2013 (13-515-516)	12	409	2% (8)	.2% (1)	6% (24)
2013 (13-517-518)	15	469	3% (12)	8% (38)	N/A
2014 (14-515-516)	12	424	4% (18)	3% (13)	N/A
2014 (14-517-518)	12	587	11% (62)	9% (53)	.3% (2)
2015 (15-515/516)	11	536	7% (39)	12% (59)	.1% (1)
2015 (15-517)	11	509	4% (21)	11% (57)	.1% (1)
2015 (15-519)	3	292	23% (36)	N/A	N/A
2016 (16-515-516)	16	431	10% (41)	3% (11)	N/A

Table 1: CTS Fingerprint Proficiency Test Results, 1995-2016

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

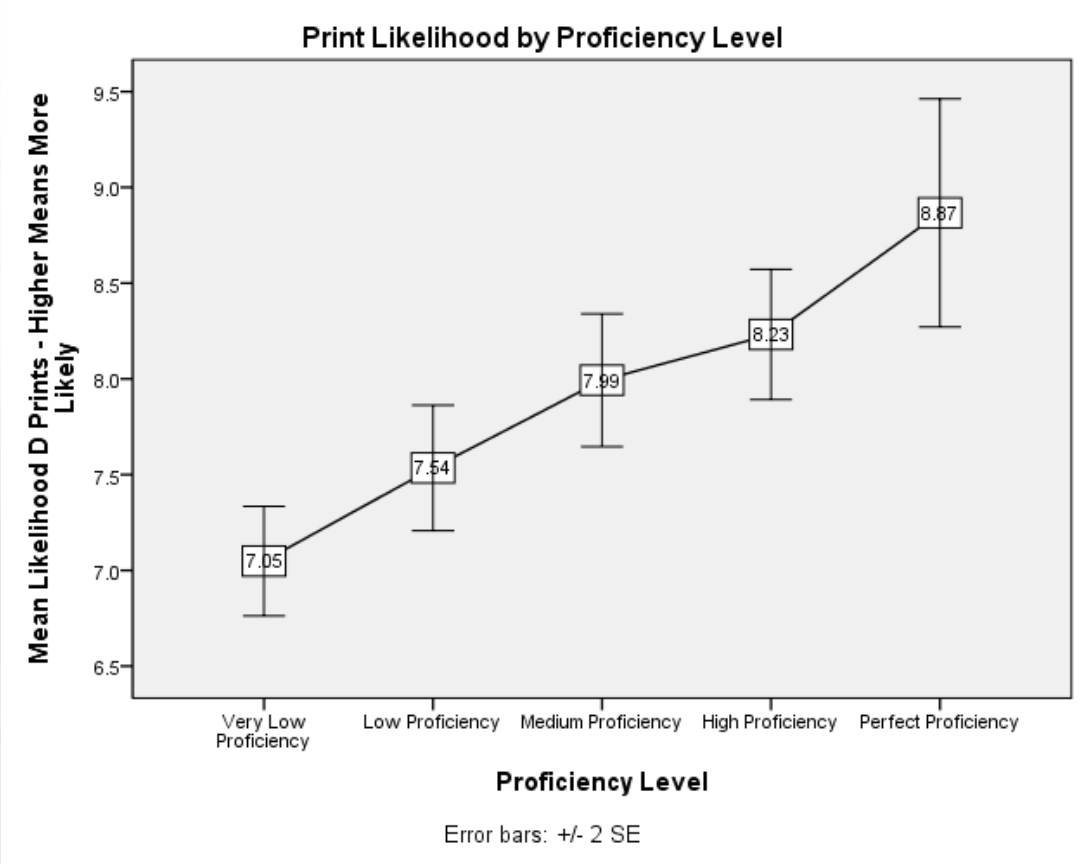
The Impact of Proficiency Testing Information on the Weight Given to Fingerprint Evidence

- Gregory Mitchell & Brandon Garrett

- We commissioned Qualtrics to recruit a nationally representative sample with respect to gender, race/ethnicity, age, income, and geographic region in the United States. A total of 1,450 adults participated in the study, which took less than 15 minutes. In addition to asking demographic questions, we gave an objective numeracy test to participants.
- The description of the case was kept simple to keep the Participants focused on the fingerprint evidence itself. The survey software assigned participants to one of 14 conditions with five proficiency levels and three error types, as well as a control in which the examiner received a perfect score on proficiency (with no errors) and a control condition with no proficiency information provided.

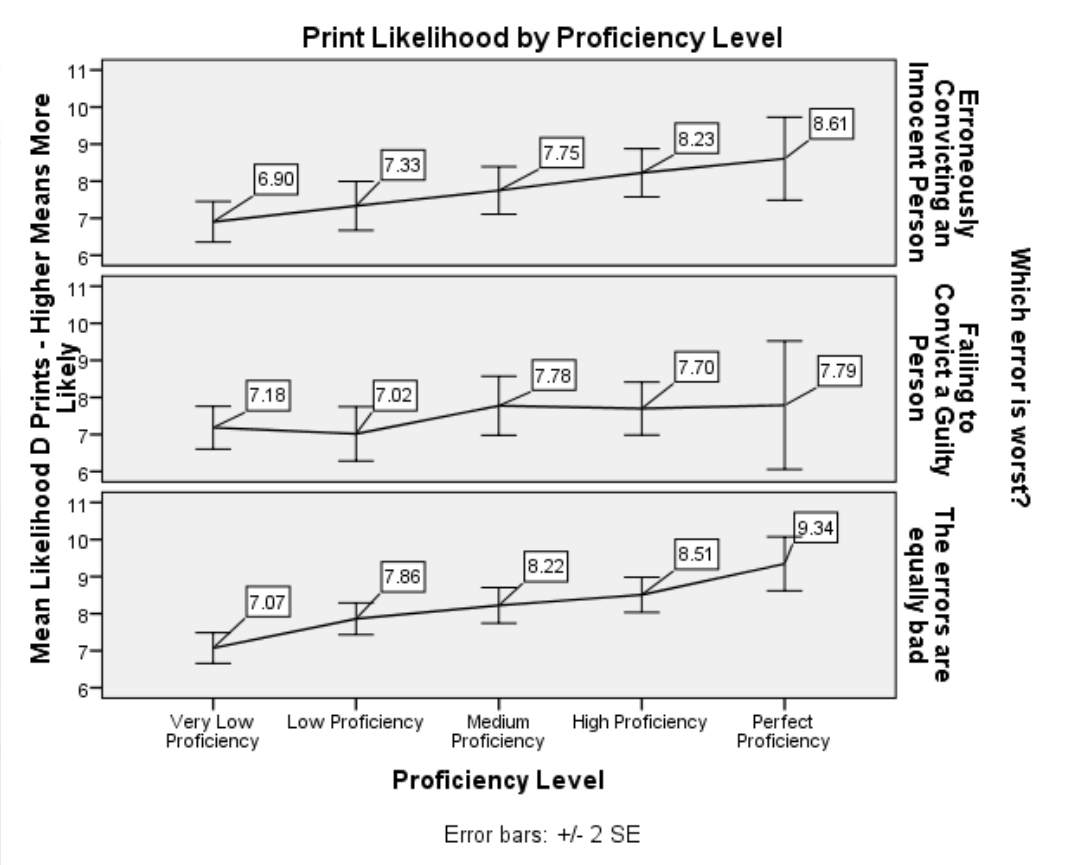
Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Print Likelihood Ratings by Proficiency Level and Error Aversion Group



Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Print Likelihood Ratings by Proficiency Level and Error Aversion Group



Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

Conclusions

- The 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 given to the examiner's identification opinion, which in turn affected judgments of the defendant's guilt.
- Those with stronger aversions to false acquittals than false convictions, older participants, and White and Asian participants gave greater weight to the fingerprint evidence, but all groups were sensitive to information about the examiner's proficiency level.
- Finally, our results suggest that jurors assume that fingerprint examiners are highly proficient but not perfect: evidence showing that an examiner's proficiency level falls below 90% is likely to inform how jurors evaluate the examiner's testimony.

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

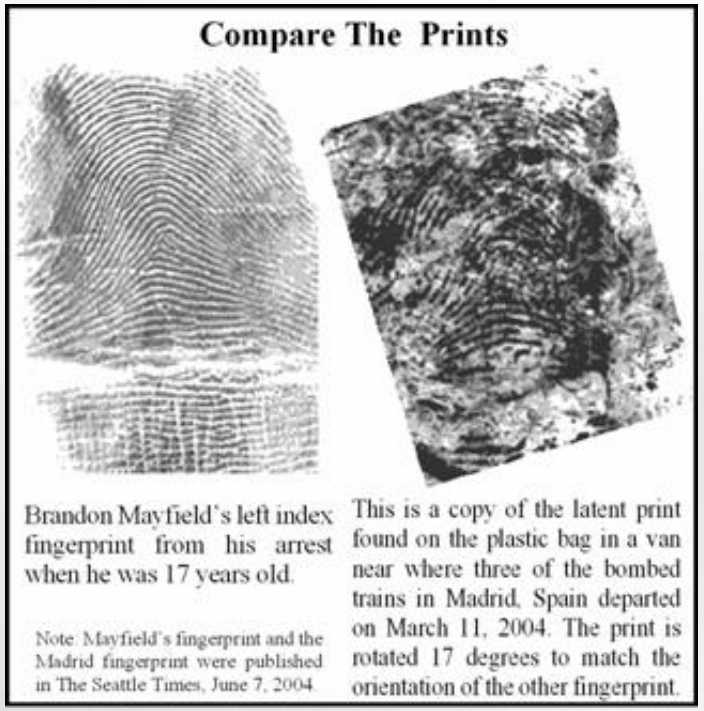
What are error rates in practice?

- PCAST Report: "We also note it is conceivable that the false-positive rate in real casework could be higher than that observed in the experimental studies, due to exposure to **potentially biasing information in the course of casework**. Introducing test samples blindly into the flow of casework could provide valuable insight about the actual error rates in casework."

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

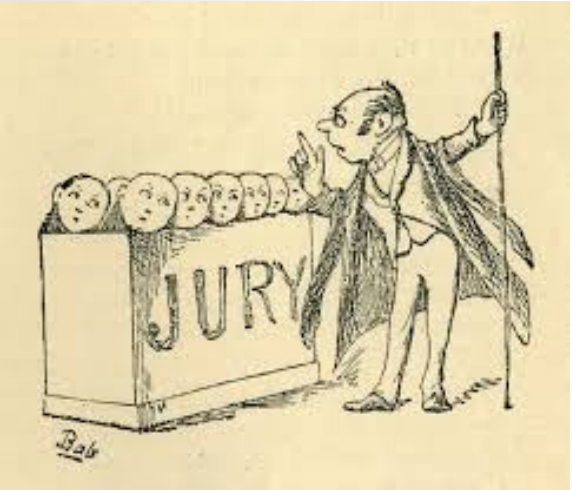
Cross on Conclusion Language

- Brandon Mayfield case (Madrid bombing):
  - Three FBI examiners gave a “100% positive identification”
  - Critical DOJ OIG Report



Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org

CLOSING



- Whether at a trial or post-conviction:
- Emphasize subjectivity in analysis
- Ambiguity of data interpreted
- Lack of standards (“black box” method)
- Presence of potentially biasing information
- Motivational influences
- Overstatement in conclusions
- False positive rates
- Proficiency data
- Show how the methods used did not safeguard against cognitive bias – and how the analysis is not error-free

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org