
William C. Thompson, J.D., Ph.D.
University of California, Irvine

Stetson University College of Law
Crime Scene to Courtroom Forensics Training Webinar
September 16, 2019
Role of AAAS
American Association for the Advancement of Science

• World’s largest scientific society
• Publisher of *Science*
• Undertook a “gap analysis” of:
  – Arson investigation
  – Latent print examination
  – Toolmark analysis
• Goals
  – Assess strengths and limitations
  – Propose a research agenda
  – Comment on appropriate reporting practices
AAAS Working Group on Latent Print Analysis

- John Black, Black & White Forensics
- Anil Jain, Dept. of Computer Science and Engineering, Michigan State University
- William Thompson, Dept of Criminology, Law & Society, University of California, Irvine
Previous reports

- 2009: Strengthening Forensic Science in the United States: A Path Forward
- 2012: Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach
- 2014: Strengthening the Forensic Sciences
- 2016: Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods
Audience Poll

1. Members of the audience: What label best describes you?
   A. Lawyer
   B. Law student
   C. Latent print examiner
   D. Another kind of forensic scientist
   E. Academic or researcher
   F. None of the above
Audience Poll

2. Are you familiar with the 2009 National Academy of Sciences report on Forensic Science?
   A. Yes
   B. No
Audience Poll

3. Are you familiar with the 2016 report of the President’s Council of Advisors on Science and Technology (PCAST) on forensic pattern-matching evidence?
   
   A. Yes
   
   B. No
4. What are the chances that a qualified examiner will mistakenly declare an “identification” when comparing fingerprints of different people?

A. Less than 1 in 1 million

B. Less than 1 in 10,000 (but more than 1 in 1 million)

C. Less than 1 in 100 (but more than 1 in 10,000)

D. Less than 1 in 10 (but more than 1 in 100)

E. More than 1 in 10
Latent Print Primer

LEVEL 1 FEATURES
- Arch
- Tented Arch
- Left Loop
- Right Loop
- Double Loop
- Whorl

LEVEL 2 FEATURES
- Line-Unit
- Line-Fragment
- Ending
- Bifurcation
- Eye
- Hook

LEVEL 3 FEATURES
- Pores
- Line Shape
- Incipient Ridges
- Creases
- Warts
- Scars
Traditional Logic of Latent Print Examination

- Every fingerprint has a unique ridge pattern
- So, if the latent has sufficient detail
- And the pattern matches the suspect
- Then the suspect must be the source of the latent print

Uniqueness + Persistence = Infallibility
--No statistics needed
Three Possible Conclusions:

- Identification (Individualization/MATCH)
- Exclusion
- Inconclusive

Uniqueness + Persistence = Infallibility

--No statistics needed
Reports and Testimony

Fingerprint examiners routinely claim to have “identified” or “individualized” an unknown mark to a single known print. This identification is often characterized as being “to the exclusion of all others” on earth to a 100 [percent] certainty, and the comparison method used is claimed to have a zero percent error rate. These claims are based on the premises that friction ridge skin is unique and permanent. (Eldridge, 2017)
Academic Critics

• Examples:
  – Simon Cole
  – Jennifer Mnookin
  – Michael Saks
  – Jonathan Koehler

• Questioning scientific foundation of latent print analysis

What could possibly go wrong?

Demise of the Theory of Discernible Uniqueness
Even if the ridge detail of every finger were unique and unchangeable, it does not follow that every impression made by every finger will always be distinguishable from every impression made by any other finger, particularly if the impressions are of poor quality (e.g., limited detail, smudged, distorted, overlaid on another impression). AAAS, p. 13
Fingerprint comparisons “are being made from one imperfect, incomplete recording to another… [hence] correctly associating a degraded mark to its true source is by no means a certainty, even were one to presume absolute uniqueness of all friction ridge skin” (Eldridge, 2017, p. 75).
“Having found as many as 10 points of unusual similarity, the FBI examiners began to ‘find’ additional features in LFP 17 [the latent print] that were not really there, but rather suggested to the examiners by features in the Mayfield prints” [Office of the Inspector General, A review of the FBI’s Handling of the Brandon Mayfield case, US Department of Justice, Washington, DC, 2006.]
Key Scientific Issues for AAAS

Is there an adequate scientific foundation for understanding:

1. The degree of variability of fingerprints among individuals?
2. The degree of variability among prints made by the same finger?
3. The accuracy of human fingerprint examiners?
4. The accuracy of Automated Fingerprint Identification System (AFIS)?
5. The potential for contextual bias and how it might be addressed?

In light of the existing scientific literature:

6. What should examiners say in reports and testimony about the value of fingerprint evidence?
Scientific Foundation: Strengths

Existing research convincingly establishes:

- A high degree of variability across prints of different fingers (between-source variability)
- A relatively low degree of variability across prints of the same finger (within-source variability)
- Hence, latent print analysis is a viable tool for human identification with substantial probative value
Scientific Foundation:
Weaknesses

• Probative value in particular cases is difficult to evaluate due to uncertainty:
  – About the rarity of particular sets of features
  – About the probability of observing particular types of variation among prints of the same person

• This creates uncertainty about when fingerprint evidence is strong enough to justify claims of “identification”
When are the features shared by a latent and exemplar rare enough that they are unlikely to be repeated in the human population?

- If < 1 in 100 billion fingers, then unlikely to be repeated
- If > 1 in 100 billion, then likely to be repeated

AAAS: “…there is no scientific evidence…that latent print examiners have the ability to estimate with the required level of precision the frequency of the feature sets observable in latent prints in the human population.” (p. 63)
AAAS on Identification

“Examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, but there is no scientific basis for estimating the number of people who could not be excluded and there are no scientific criteria for determining when the pool of possible sources is limited to a single person.” (AAAS, p. 6)
PCAST on Need to Test Accuracy

Scientific validity and reliability require that a method has been subjected to empirical testing, under conditions appropriate to its intended use, that provides valid estimates of how often the method reaches an incorrect conclusion…. Nothing—not training, personal experience nor professional practices—can substitute for adequate empirical demonstration of accuracy. (PCAST, 2016, p. 46)

BJA Comment: The PCAST information presented is historical in nature. The PCAST Report recommendations were not adopted by the U.S. Department of Justice.

Presenter’s Comment: The scientific analysis in the PCAST report is relevant to present and future, not just to past, regardless of whether DoJ accepts PCAST’s policy recommendations.
AAAS Review: Accuracy of Latent Print Examiners

• When trained examiners compare known-source prints in “black-box” studies:
  – There are relatively few false identifications,
  – But higher numbers of false exclusions.

• Error rates in actual practice:
  – May be higher or lower than observed in studies
  – Likely vary depending on such factors as:
    • The quality of the prints
    • The quantity of ridge detail present
    • Whether the known print was selected for comparison based on similarity to the latent
Ulery et al. (2011)—aka FBI Black-box Study

– 169 examiners each compared about 100 print-pairs
– 23% of comparisons were “no value”
– 29% were “inconclusive”
– 4083 non-mated pairs
  • 3638 conclusive calls
  • 6 false ID’s by 5 examiners (0.17%)
    – PCAST puts upper 95% confidence bound at 0.33%, or 1 error in 306 non-mated comparisons
– For mated pairs, false exclusion rate was 7.5%
– 85% of participants had a false exclusion

BJA Comment: The PCAST information presented is historical in nature. The PCAST Report recommendations were not adopted by the U.S. Department of Justice.

Presenter’s Comment: Neither the passage of time, nor changes in DoJ policy, has affected the way in which confidence intervals are calculated.
Difficult Non-Mated Pairs—Ulery et al. (2011)
Pacheco et al. (Miami-Dade Study)(2014)

• 109 examiners evaluated 40 latents
• Compared to 10-prints refs from 3 individuals
• For comparisons on which they made a call:
  – False positive rate 4.2%; false negative rate 8.7%
  • PCAST 95% upper bound for false positives—5.4%, or 1 false match in 18 comparisons
  – Excluding possible “clerical errors,” false positive rate 0.7% (95% upper bound 1.5%)

BJA Comment: The PCAST information presented is historical in nature. The PCAST Report recommendations were not adopted by the U.S. Department of Justice.

Presenter’s Comment: History has nothing to do with the calculation of confidence intervals; neither does the DoJ’s position on PCAST’s policy recommendations.
Pacheco et al. (Miami-Dade Study)(2014) (con’t.)

• Reference prints NOT chosen for similarity with latents
• All false positives and half of false negatives were “detected” in a “verification” stage
  – AAAS: “not clear whether verification would be as effective in practice as observed here.” (p. 57)
• Not yet published
• More research clearly needed
Liu, Champod & Luo (2015)

• 40 Chinese examiners
• 5 “difficult cases”—non-mated pairs from database
• 1 “close non-match” (CNM) produced a false positive in 3 of 27 examiners who found it to be “of value”
• “as database size is increasing, examiners’ experience alone is no longer sufficient to deal with CNM prints”
• Will we be seeing more cases like Mayfield?
AAAS on Accuracy

• Additional research is needed to evaluate factors that affect the performance of latent print examiners
• The research should be done by introducing known-source research samples into the routine flow of casework in a manner that makes them indistinguishable from casework evidence
• Government agencies should take steps to facilitate such research (as called for by the National Commission on Forensic Science)
Recommendations of the National Commission on Forensic Science (2016)
Remaining Challenges

- Why are there so many false exclusions?
- Is it possible that examiners do not fully understand the degree of within-source variability?
Difficult Mated Pairs

Pair A:
- 2 Exclusions
- 2 No value
- 8 Inconclusive
- 2 Individualizations

Pair B:
- 5 Exclusions
- 2 No value

(Ulery et al., 2011)
Black Box Study Respondents

25. Are you aware of ever having made an erroneous conclusion (after training)? (Check all that apply - may add to over 100%)

<table>
<thead>
<tr>
<th>Response</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>No response</td>
<td>2</td>
<td>1%</td>
</tr>
<tr>
<td>No</td>
<td>103</td>
<td>65%</td>
</tr>
<tr>
<td>Yes, on casework; detected after it was reported to a contributor</td>
<td>10</td>
<td>6%</td>
</tr>
<tr>
<td>Yes, on proficiency test only</td>
<td>4</td>
<td>3%</td>
</tr>
<tr>
<td>Yes, on casework; detected during verification</td>
<td>43</td>
<td>27%</td>
</tr>
</tbody>
</table>

On question 25, one examiner indicated yes both on a proficiency test and on casework detected during verification. Two examiners indicated yes both on casework detected after reporting and on casework detected during verification.
Black Box Study Respondents

<table>
<thead>
<tr>
<th>Description</th>
<th>Cases</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any comparison that is not an individualization is an exclusion</td>
<td>7</td>
<td>4%</td>
</tr>
<tr>
<td>Exclusion means that the latent did not come from any finger for that subject but could have come from other friction ridge skin (e.g., palm) from that subject</td>
<td>16</td>
<td>10%</td>
</tr>
<tr>
<td>Exclusion means that the latent did not come from any friction ridge skin for that subject</td>
<td>81</td>
<td>51%</td>
</tr>
<tr>
<td>Exclusion means that the latent did not come from the source of the exemplar (e.g., a specific finger), but could have come from another finger from that subject</td>
<td>18</td>
<td>11%</td>
</tr>
<tr>
<td>Not used</td>
<td>37</td>
<td>23%</td>
</tr>
</tbody>
</table>
Differing Policies on “Exclusion”

The participants in these studies came from agencies with differing policies with respect to whether and how exclusions are used, whether exclusions are verified, whether examiners are discouraged from making inconclusive decisions, and how latents of value for exclusion only should be treated. Some of the erroneous exclusions may be due to lack of familiarity with the concept of exclusion: some examiners apparently confuse exclusions and non-identifications. Standardization of exclusion terminology, policies, and procedures is needed. (Ulery et al., 2017).
AFIS

• Accuracy is high
• AFIS are enormously useful for screening
• Greater transparency regarding operation of proprietary systems would help advance the science
• AFIS do not currently provide useful estimates of weight of evidence, but can could evolve over time to do so
  – More on this later…
Contextual Bias

- Latent print examiners are vulnerable to contextual bias
  - Investigative facts
  - Exposure to reference print
- Bias can occur without conscious awareness and cannot reliably be suppressed by the individual examiner
- Bias can be mitigated through the use of context management procedures
  - Which have been adopted by some laboratories
  - And should be adopted more generally
Preliminary communication

Contextual information renders experts vulnerable to making erroneous identifications

Itiel E. Dror *, David Charlton, Ailsa E. Péron

School of Psychology, Faculty of Medicine, Health and Life Sciences, University of Southampton, Southampton SO17 1BJ, UK

Received 1 June 2005; received in revised form 17 October 2005; accepted 17 October 2005
Available online 1 December 2005

Abstract

We investigated whether experts can objectively focus on feature information in fingerprints without being misled by extraneous information, such as context. We took fingerprints that have previously been examined and assessed by latent print experts to make positive identification of suspects. Then we presented these same fingerprints again, to the same experts, but gave a context that suggested that they were a no-match, and hence the suspects could not be identified. Within this new context, most of the fingerprint experts made different judgements, thus contradicting their own previous identification decisions. Cognitive aspects involved in biometric identification can explain why experts are vulnerable to make erroneous identifications.

© 2005 Elsevier Ireland Ltd. All rights reserved.

Keywords: Psychology; Cognition; Erroneous identification; Bias; Extraneous information; Contextual influence; Fingerprints
Only one participant (20%) judged the prints to be a match, thus making a consistent identification regardless of extraneous context. The other four participants (80%) changed their identification decision from the original decision they themselves had made five years earlier. Three of these four participants directly contradicted their previous decision and now judged the fingerprints as definite non-matches, whereas the fourth participant now judged that there was insufficient information to make a definite decision (either a match or a non-match. (Fig.2).
Context Management Procedures

- Case Management/Blinding
- Sequential Processing
  - Linear ACE-V
    - Analysis, Comparison, Evaluation, Verification
  - Linear Sequential Unmasking
Context management procedures ... are necessary to protect the scientific integrity of latent print evidence. .... Failure to control for contextual bias (when it is possible to do so) is unacceptable in the broader scientific community. It should be unacceptable in forensic science as well. (AAAS, p. 42).
Recommendations of the National Commission on Forensic Science (2015)
What Should Examiners Say in Reports and Testimony

• Examiners should not claim that they can narrow the potential sources of a latent print to a single finger.

• But what are the alternatives?
  – “Identification” with Caveats
  – Quantitative/Statistical Statement
  – Qualitative Statement About Source Probability
  – Qualitative Statement About Strength of Evidence
“Identification” with Caveats

• PCAST: Error rates observed in black-box studies should be reported.

• DOJ Uniform Language: Examiners should avoid:
  – Using phrase “to exclusion of all others”
  – Implying absolute or numerically calculated certainty
  – Implying that error rate is zero

• AAAS:
  – Avoid term “identification”
  – Be prepared to report error rates
  – Three proposed caveats

BJA Comment: The PCAST information presented is historical in nature. The PCAST Report recommendations were not adopted by the U.S. Department of Justice.

Presenter Comment: AAAS agrees with PCAST that error rates found in research should be reported. DoJ cannot prevent inquiry on this issue during trial.
Latent print examination allows examiners to draw conclusions about whether two friction ridge impressions could have originated from the same source. These conclusions are opinions, they are not facts.
It is not possible for a latent print examiner to determine that two friction ridge prints originated from the same source to the absolute exclusion of all other sources. A latent print examiner may be able to exclude a substantial proportion of the human population as the source of a latent print, but it is not possible to determine how many people would not be excluded, nor is it possible to determine when the pool of possible sources is limited to a single person.
Studies have shown that latent print examiners are highly accurate in associating latent prints with reference prints known to be from the same source, and in excluding reference prints known to be from a different source. But latent print examination is not infallible. Both false associations and false exclusions have occurred in studies of examiner performance and in actual cases.
Quantitative/Statistical Reports

• It is not possible to compute random match probabilities (like those reported in some DNA cases)

• But statistical models are being developed and show great potential
  – *E.g.*, US Army Lab’s FR STAT
  – Similarity scores for suspect vs. database
  – Avoid problem of contextual bias
  – Results are strong (but not definitive)
A method for the statistical interpretation of friction ridge skin impression evidence: Method development and validation

H.J. Swofford\textsuperscript{a,}\textsuperscript{*}, A.J. Koertner\textsuperscript{a}, F. Zemp\textsuperscript{b}, M. Ausdemore\textsuperscript{c}, A. Liu\textsuperscript{d}, M.J. Salyards\textsuperscript{a}

\textsuperscript{a}U.S. Army Criminal Investigation Laboratory, Defense Forensic Science Center, USA
\textsuperscript{b}School of Criminal Justice, Forensic Science Institute, University of Lausanne, Switzerland
\textsuperscript{c}Department of Mathematics and Statistics, University of South Dakota, USA
\textsuperscript{d}Department of Statistics, University of Virginia, USA
Reporting Source Probabilities

- Why not say the suspect is “probably the source” or “very probably the source”?
  - Acknowledges uncertainty
  - While also recognizing strength of evidence
- The Elvis Conundrum
Elvis Conundrum

• A latent print from a crime scene has a friction ridge pattern indistinguishable from that of Elvis Presley

• How rare does the pattern need to be to prove Elvis:
  – Is more likely than not to be the perpetrator?
  – Is highly likely to be the perpetrator?
  – Can be “identified” as the perpetrator?
Conundrum

- Any statement about the probability Elvis is the source depends on the examiner’s assessment of the strength of Elvis’ alibi
- But this is true for every suspect, not just Elvis
- Should latent print examiners be evaluating the strength of suspects’ alibis when deciding what to say about their findings?
Reporting Strength of Evidence

• Army Defense Forensic Science Center
  Reporting Language:
  “the likelihood of observing this amount of correspondence when impressions are made by different sources is considered extremely low”

• European Network of Forensic Science Institutes (ENFSI)—Likelihood ratios
  “the level of correspondence observed is X times more probable if the suspect is the source of the print than if the print came from a random person”
Reporting Strength of Evidence

• “In my opinion, the level of correspondence observed is far more probable if the suspect is the source of the print than if the print came from a random person”

• “and this provides extremely strong support for the hypothesis that the suspect is the source”
More on Reporting Methods


More on Lay Reactions to Alternative Reporting Methods

Any Questions????

William Thompson: william.thompson@uci.edu
John Black: john@bwforensics.com
Thank you very much

Email: william.thompson@uci.edu