2018

Fingerprint Science

Joseph B. Kadane
Carnegie Mellon University

Abstract

This paper examines the extent to which data support the source attributions made by fingerprint examiners. It challenges the assumption that each person's fingerprints are unique, but finds that evidence of persistence of an individual's fingerprints is better founded. The use of the AFIS (Automatic Fingerprint Identification System) is problematic, because the algorithms used are proprietary. Additionally, the databases used in conjunction with AFIS are incomplete and not public. Finally, and most crucially, the finding of similarities between the mark found at a crime scene and a fingerprint on file does not permit estimation of the number of persons in a given population who share those characteristics. Consequently, there is no scientific basis for a source attribution; whether phrased as a "match," as "individualization" or otherwise.

Keywords

Fingerprint uniqueness, fingerprint persistence, AIS, source attribution, individualization, match

Disciplines

Forensic Science and Technology

Comments

This article is published as Kadane, Joseph B. "Fingerprint science." *The Annals of Applied Statistics* 12, no. 2 (2018): 771-787. Posted with permission of CSAFE.

The Annals of Applied Statistics
2018, Vol. 12, No. 2, 771–787
https://doi.org/10.1214/18-AOAS1140

FINGERPRINT SCIENCE

BY JOSEPH B. KADANE

Carnegie Mellon University

This paper examines the extent to which data support the source attributions made by fingerprint examiners. It challenges the assumption that each person's fingerprints are unique, but finds that evidence of persistence of an individual's fingerprints is better founded. The use of the AFIS (Automatic Fingerprint Identification System) is problematic, because the algorithms used are proprietary. Additionally, the databases used in conjunction with AFIS are incomplete and not public. Finally, and most crucially, the finding of similarities between the mark found at a crime scene and a fingerprint on file does not permit estimation of the number of persons in a given population who share those characteristics. Consequently, there is no scientific basis for a source attribution; whether phrased as a "match," as "individualization" or otherwise.

1. Contributions of Steve Fienberg to the nexus between statistics and forensic science. Soon after Steve Fienberg joined the Statistics Department at Carnegie Mellon in 1980, Steve, Morrie Degroot and I recognized that we all had served as expert witnesses. We saw the law as a fruitful area of statistical application to explore. This led us to our jointly edited book, DeGroot, Fienberg and Kadane (1983). It also led to the review paper of Fienberg and Straf (1982), announcing the establishment of an NRC panel on Statistical Assessments as Evidence in the Courts. This panel was jointly sponsored by two committees of the National Research Council: the Committee on National Statistics and the Committee on Research on Law Enforcement and the Administration of Justice. The panel's report [Fienberg (1989)] concentrated on civil cases, such as discrimination by age, race or sex, and commercial disputes.

Later, attention shifted to the criminal law. Steve chaired the Committee to Review the Scientific Evidence on the Polygraph (2003). A parallel Committee on Scientific Assessmant of Bullet Lead Elemental Composition Comparison (2004) was also conducted. These reports showed that the methods used in these two forensic disciplines were not scientifically based, and led to (some) reforms. In the same period, Steve also led the Sackler Conference on Forensic Science: The nexus of science and law [Fienberg (2005)]. The NAS/NRC panel on forensic science [Committee on Assessing the Needs of the Forensic Science Community (2009)], led to increased attention to the parlous state of science more generally

Received July 2017; revised November 2017.
Key words and phrases. Fingerprint uniqueness, fingerprint persistence, AIS, source attribution, individualization, match.

in the forensic sciences. In response, the Obama Administration appointed a National Commission on Forensic Science, of which Fienberg was the sole member who was a statistician. Fienberg was also very active in putting together the proposal, with Alicia Carriquiry, Bill Eddy, Karen Kafadar and Hal Stern, that led to the founding and funding of CSAFE (Center for Statistics and Applications in Forensic Evidence). In that connection, Steve asked me to look into fingerprints. The remainder of this paper is a summary of what I found.

2. Several cultures. The forensic sciences live at the crossroads of several rather different cultures. First, they are aimed at helping the legal community by advising on the import of various kinds of evidence. Historically they have been funded by, and usually report to, the prosecution or police in criminal cases. (I'll come back later to this point to discuss the extent to which this history and current practice may lead to bias in their work.)

The second culture that affects the forensic sciences is science itself. Science tends to be cautious about making claims, and careful about conclusions. The well-known 2009 National Academy of Sciences report, "Strengthening the Forensic Sciences in the United States: A Path Forward" [National Research Council (2009)], on the forensic sciences criticized many forensic disciplines for their lack of careful experimental work to support their claims.

The third culture is the forensic community itself, which has its own journals, professional societies and traditions. Forensic laboratories concentrate on case work, and are often underfunded and overworked. They interact with the legal and scientific communities in various ways, some comfortable, some not.

Among the forensic science disciplines, fingerprint analysis holds a special place. Until the advent of DNA, it was regarded as the best founded forensic discipline. It is very widely used in criminal cases, and has been the subject of recent scrutiny. This paper focuses on fingerprint analysis.

3. Early recent history of fingerprint analysis. While there were precedents for interest in fingerprints dating back centuries, fingerprint analysis started to be taken seriously as a way to identify people with the publication of two letters in the English journal *Nature* in 1880. Henry Fauld (1880) wrote the first letter, published on October 8. He had been a missionary in Japan, and had gotten interested in finger marks on pottery, and later studied fingerprints of various species of monkeys. He moved on to people, and recounts two criminal cases in which he found fingerprints useful. The publication of Fauld's letter led to a response from William Herschel (1880), who had been an administrator in British India. He had been using fingerprints to certify the identity of signers of documents. (Later these two would contest who deserves the greater credit for suggesting the use of fingerprinting to identify people.)

Soon after this exchange, Mark Twain (Samuel Clemens) published a short story, "A Thumb-print and What Came of It," in 1883. It's a wonderful tale, the crux of which is that a murderer was identified by a bloody thumbprint left at a crime scene. This story was published as part of Twain's "Life on the Mississippi."

The first scientific investigation of fingerprints was undertaken by the statistician and geneticist Francis Galton (1888, 1892), first as a letter in *Nature* in 1888, and then in his book, "Finger Prints," published in 1892.

Twain's "The Tragedy of Pudd'nhead Wilson" (1894) revolves around children switched at birth, but unmasked later by fingerprints collected at birth. Conan Doyle's Sherlock Holmes finds in "The Norwood Builder" (1903) that a fingerprint had been added to a crime scene, which enables him to detect a deception aimed at convicting an innocent man.

The importance of these fictional uses of fingerprints is that they introduced the technique and its possible consequences to the nonprofessional public, laying the groundwork in popular culture for its use in court.

4. What fingerprint analysts do. What follows below is a very brief exposition of the main steps taken in fingerprint analysis. The Fingerprint Sourcebook [National Institute of Justice (2011)] gives a much fuller description.

A fingerprint analyst is presented with a mark found at a crime scene. The analyst first determines whether the mark is of such quality that there is reasonable hope of analysis. (Some aren't, if they were smudged or degraded, for example.) If the mark is regarded as having adequate quality, the mark is then put through an AFIS system (Automated Fingerprint Identification System) against a database of prints of known persons. AFIS identifies a handful of prints judged most similar to the mark. The analyst then compares the mark to each such print on the basis of structure and minutiae, and makes a "decision." This work is then (supposed to be) verified by an independent second examiner [National Institute of Justice (2011)]. When the verifier knows the outcome of the original analysis, as is often the case in the U.S., the independence of the verification is impugned. If verification is requested only for identification conclusions, the verifier already knows the analyst's opinion. For a system in which independence is maintained, see Mustonen, Hakkarainen and Tuuainen (2015).

Fingerprint analysts often point to the acronym ACE-V to explain what they do. The letters stand for Analysis, Comparison, Evaluation, and Verification. However what each of these activities actually comprise is not uniquely established, so two different analysts can perform different activities, and arrive at different conclusions, while both can claim to be using the ACE-V methodology. In this sense, it is more an outline than a scientific method. Some commentary can be found in Haber and Haber (2008) and Langenburg (2012).

5. Questions and issues.

a. Uniqueness

There are two senses of what might be meant by uniqueness. The first is an ideal representation of the finger with no loss of information. However, latent prints collected at a crime scene are generally partial, and can be smudged or otherwise degraded. The first sense is typically the sense used in discussions of uniqueness, and is the sense intended here.

There are two basic approaches to the question of whether a person's fingerprint on a specified finger could be the same as another person's fingerprint. The first is whether there is a basis in biological theory to support such a claim. For example, there is such theory to claim that DNA of fraternal twins are distinct, relying mainly on Mendelian genetics and the enormous number of genes and nongenetic loci in DNA. An important series of studies [Srihari, Srinivasan and Fang (2008), Liu and Srihari (2009)] shows that identical twins have patterns of fingerprints that are more similar to each other than fraternal twins, and both have patterns of fingerprints more similar to each other than do unrelated persons. These results support the conclusion that fingerprints are determined only partially by genetics. However, the biology of the development of fingerprints in the fetus does not address the question of whether such development guarantees uniqueness [Maceo (2009)].

A second source of information about uniqueness is empirical. At present, "no two people in the world have been found to have matching fingerprints" [Langenburg (2011)]. To what extent does this empirical finding support the proposition of uniqueness? According to the U.S. Census Bureau there are more than 7.5 billion people in the world [United States Census Bureau (2017)]. Therefore, there are $(7.5 \times 10^9)^2/2 = 2.8 \times 10^{19}$ pairs of people, and since each of them has 10 fingers, there are 100 comparisons to make. Therefore, there are 2.8×10^{21} comparisons involved in the claim of uniqueness. This is an impressively large number. For example, fingerprint analysis in its current form started with Galton (1892), say roughly 130 years ago. Since that time the number of seconds that have elapsed is $(60)^2(24)(365.25)(130) = 4.1 \times 10^9$. Therefore, it would require over 6.8×10^{11}, that is, 680 billion, unique comparisons every second since 1892 to verify the hypothesis empirically. Furthermore, people are born and die every day. It is hard to believe that any but a tiny fraction of the necessary comparisons have been made. Consequently, the empirical claim that no two people with identical prints have been found is distinctly unimpressive.

This view is not unanimously held. For example, the International Association for Identification (2007) wrote "The IAI fully supports the principle that finger, palm and footprints (friction ridge detail) are unique to each and every individual. This principle has been well established through the biological sciences of anatomy, embryology and genetics [...] As yet, no two fingerprints from different individuals have ever been found to be the same." This statement is peculiar in that it cites no references from the sciences mentioned to support uniqueness. Furthermore, if these sciences have "well established" uniqueness, why bother with the weak empirical evidence?

b. Permanence

The research on persistence is more reassuring. By comparing the fingerprints of 15,597 persons arrested in Michigan from five to 12 years apart, Yoon and Jain (2015) find that fingerprints do recognizably persist.

Informal discussion suggests that people who do a lot of work with their hands gradually wear down their fingerprints. Their fingerprints haven't changed, but they can be less distinct, making them harder to recover at a crime scene or to print under controlled circumstances. Additionally injuries and perhaps certain drugs can alter the appearance of fingerprints. Nonetheless, in the main, fingerprints do seem to be remarkably stable over time.

c. Infallibility

For a long time, fingerprint analysis was touted as infallible. For example, in 1985, the FBI's "The Science of Fingerprints: Classification and Uses," declares "Of all the methods of identification, fingerprinting alone has proved to be [...] infallible." Nonetheless, there is now a long string of cases in which fingerprint analyses have been proven to be mistaken. Cole (2005) lists 22 of them, the most famous of which is the false accusation against Oregon attorney Brandon Mayfield. He was wrongly accused of participating in the Madrid bombing on the basis of a fingerprint identification certified by three of the FBI's most senior fingerprint analysts, and confirmed by an independent fingerprint analyst hired by the defense [Department of Justice (2006)]. See also Zabell (2005). Another important case is that of Shirley McKie, a Scottish police officer wrongly accused of perjury for testifying that she had never set foot in a particular crime scene; the police believed they had found her fingerprint in that house. The police were very reluctant to confront the possibility that they had made an error [Specter (2002)].

Additionally, there is evidence that fingerprint experts differ in their assessments of the same print, and even differ with themselves when presented with the same prints at a later time [Dror et al. (2011)].

In view of this litany of erroneous identifications, it seems absurd to keep defending fingerprint evidence as infallible. The claim of infallibility, however, fit with a broader effort by the FBI under J. Edgar Hoover to promote the FBI in specific, and law enforcement in general, as smarter, more scientific and more capable than the facts warrant [Cecil (2014)].

In 1979, the International Association for Identification, in resolution 1979-7 deemed "testimony of possible, probable, or likely friction ridge identification" to be "conduct unbecoming." In 1980, this resolution was clarified, in resolution 1980-5, to require "a statement that the print could be that of someone else." In turn, both resolutions were rescinded by resolution 2010-18, which stated that "The practice (of friction ridge comparison) by trained and competent examiners has been shown, through experience and study, to be reliable with rare occurrences of error."

However, actual practice does not necessarily follow any of these prescriptions. As late as 2015 in a case in Florida, a fingerprint expert was testifying to 100%

certainty of an identification against a juvenile. In this case, the only evidence was a fingerprint (Florida v. S. Hayes). Langenburg (2012), page 65, proposes that the origin of forensic sciences in support of prosecutions, and the adversarial system of US courts, both encourage statements of exaggerated certainty with respect to fingerprint conclusions.

d. AFIS (Automatic Fingerprint Identification System)

I start this discussion by making two bold assumptions. (After discussing the implications of these assumptions, I address the extent to which the real world differs from these assumptions, and what consequences ensue.)

Assumption A. *The AFIS system is unerringly accurate in finding that member of the database of fingerprints most similar to the mark found at the crime scene, and returns real-valued similarities without ties.*

This assumption leaves unexamined (for the moment) the question of what is meant by "similar;" the intent of the assumption is to give the benefit of the ambiguity to the AFIS system.

Assumption B. *The database of fingerprints used by the AFIS system is a random sample of the relevant population's fingerprints.*

Suppose the database has fingerprints of n people, with similarities to the mark $x_1 < x_2 < \cdots < x_n$, so x_n is the most similar to the mark found at the scene of a crime. Under these assumptions only the fingerprint most similar to the mark (with similarity x_n) is a possible match. Its probability of being correct is n/N, where N is the size of the population. To put some numbers to this, compare a national database of 116 million to a local database of 10 thousand. The population might be taken to be either the US population (324 million) or the world population (7.5 billion). Table 1 records the results.

Thus, the only case with appreciable probability of finding the correct match is the national database and the national population, and then it is only 35.8%.

TABLE 1
Probability that the person whose fingerprint has similarity x_n is the culprit under Assumptions A and B

Description	N	n	**Probability that match is correct**
World comparison national database	7.5 billion	116 million	0.015
World comparison local database	7.5 billion	10,000	1.3×10^{-6}
US population national database	324 million	116 million	0.358
US population local database	324 million	10,000	3.09×10^{-5}

As promised, I now consider relaxing Assumption A made above. In fact, there are several AFIS systems, and their results are not identical (therefore they can't all be perfect) [Moses (2011)]. Also, they are proprietary, so no one outside the company that owns it knows the algorithm being used. Additionally, they are applied to a variety of different databases.

Imagine that an analyst chooses to examine the 10 most similar fingerprints in a database. This opens the fingerprint analyst to the following hypothetical cross-examination:

Q: Do I understand correctly that you examined the fingerprints of only 10 people in the database you used?
A: Yes.
Q: Why did you not examine the fingerprints of the other people in the database?
A: Because the agency requests AFIS output of the 10 fingerprints most similar to the mark found at the crime scene.
Q: Do you know what measure of similarity is used in the AFIS system you employed?
A: I believe that AFIS similarities measure the physical distance and angle between the marked minutiae in both the searched and the candidate prints. Generally, greater similarities increase the score.
Q: Let me ask again. Do you know the formula used to calculate the similarity measure used by the AFIS system you employed?
A: No.
Q: Then what expertise do you have to offer the court on whether other people in the database might have fingerprints more similar to the mark found at the crime scene than those of the defendant?

Now reconsider Assumption B, that the database consists of the fingerprints of a random sample of people either in the world or in the US population. Of course, it doesn't. Fingerprints are taken of people generally because they have been arrested, or because they hold particularly sensitive jobs. A worldwide random sample could include, for example, a Mongolian cattle-herder, someone very unlikely to have committed a crime in the U.S. Therefore, common sense suggests that local or national databases are more likely to contain the fingerprints of a culprit than a random world-wide database. The question of "how much more likely" is outside of the expertise of a fingerprint analyst, as the following hypothetical cross-examination shows:

Q: Please describe your training as a fingerprint examiner.
A: I was taught how to compare two fingerprints, taking care to identify minutiae in both, and how to account for possible distortion depending on how the mark may have been placed on the object where it was found.
Q: Do you agree that the question of whether the person who left this mark at the crime scene is in the database is an issue of statistics, criminology and the sociology of crime?

A: Yes.
Q: Were you trained in the sociology of crime, criminology and statistics?
A: No.
Q: So you are not in a position to opine, on the basis of your expertise in fingerprint analysis, on the probability that the person who left this mark is in the AFIS database utilized in this case?
A: No.

It should also be pointed out that data on the proportion of crimes committed by persons who live close to the scene of the crime are likely to be biased. Police tend to look locally, so disproportionately the unsolved crimes may be committed by nonlocals.

e. Contexual bias

Fingerprint examiners are after all human, and hence subject to the cognitive limitations we all have. As described by Dror et al. (2011),

> "The human cognitive system is limited in its capacity to process information. The information available far exceeds available brain power and cognitive resources, and therefore we can only process a fraction of the information presented to us. This mismatch between computational demands and available cognitive resources caused the development of cognitive mechanisms that underpin intelligence. For example, we prioritize what information to process according to our expectations [...] Expectations are derived from experience, motivation, context, and other top-down cognitive processes that guide visual search, allocation of attention, filtering of information, and what (and how) information is processed. These mechanisms are vital for cognitive processes to be successful. Expertise is characterized by further development and enhancement of such mechanisms [...]
> Therefore, there are good scientific data showing that the presence of any contextual information may affect cognitive information processing. Various factors and specific parameters define the context, whom it may affect, how, and to what extent."

The particular contextual information at issue with respect to fingerprint analysis is what information other than the prints themselves should be made available to the analyst at the time of the examination. It seems obvious that some additional information is warranted, such as the nature of the surface from which the mark at the crime-scene was lifted. But should the analyst be made aware of the nature of the crime in question, the other evidence against the suspect, etc.?

In one dramatic experiment, by Dror, Charlton and Peron (2006), five experienced examiners were each presented with fingerprints that they had examined before, and assessed as a positive identification. The examiners were told (falsely) that these were prints from the (infamous) Brandon Mayfield case. The examiners were asked to examine the prints, paying attention only to the prints. Three of the five changed their decision to exclude the suspect they had previously positively identified. A fourth changed from positive identification to undetermined. Thus the contextual information had a strong effect. Later work of Dror and Charlton (2006) and Dror and Rosenthal (2008) confirmed this effect with less drastic contextual information.

It might be argued that, in analogy to medical diagnosis, a fingerprint analyst might come to more accurate conclusions the more information he or she knows about the case. For medical diagnosis, this argument makes a lot of sense. However, the legal context requires a different way of thinking. A fingerprint expert testifying in court is there to give an opinion about fingerprints, not about the case as a whole. To move beyond fingerprints is to usurp the role of the factfinders, the jury or the judge. Thus the analyst's information should be narrowed only to what is necessary to appreciate the fingerprints themselves, independent of the ramifications the fingerprint conclusions may have on the rest of the legal and factual issues before the court.

f. Proficiency examinations

The National Academy Report [National Research Council (2009)] recommended developing proficiency testing in all forensic disciplines, including fingerprint analysis. In order to see why this is difficult, consider the following simplified scenario of testing.

Imagine teaching a mathematics class, and giving a test. One can give a "hard" test, in which the weaker students will perhaps perform uniformly poorly, while the stronger students will have the opportunity to differentiate themselves from each other. Alternatively, one could give an "easy" test, which can result in the stronger students results being very similar, but the weaker students can differentiate themselves from each other. In both cases, the instructor can learn which questions were harder than others, and which students are stronger than others. I refer to results of this kind as "comparative" results, as the comparison is to others (questions or students) within the examination. However the class average is a function of both the class and the test, and therefore not in itself particularly meaningful. I refer to results of this kind as "absolute" results, as it invites comparisons outside the examination. Hence we may expect to be able to compare certain groups of test-takers to others, or certain groups of test items to others. But we should not expect to be able to assess the overall performance of the class, nor the overall difficulty of the examination.

Now suppose that an observer is invited to use the outcome from this mathematics exam to judge the probability that a specified person will correctly answer a given mathematics question. The observer would have to consider how the mathematical skill of the person compares to the skill of those of the exam-takers. Indeed there are likely to be a variety of different skills at play in solving a mathematics problem, as some math solvers think more geometrically, while others think more algebraically, for example. The observer would also have to consider how the difficulty of the mathematics question compares to those on the exam, and again "difficulty" may have several important dimensions. Both of these issues would require the observer to have considerable knowledge and skill in order to have a reasoned judgment about the relevance of the exam results.

Testing fingerprint analysts has all these issues, and more. Responding to the call of the National Academy report, in a landmark study, Ulery et al. (2011), tested 169 latent print examiners who each saw approximately 100 pairs of prints from a pool of 744 pairs, of which 520 were true mates from the same finger, and 224 were not. The paper reports that "There is substantial variability in the attributes of latent prints, in the capacities of latent print examiners, in the types of casework received by agencies, and the procedures used among agencies." (ibid., page 7734). The examiners in this study were aware that they were being examined. In the mathematics scenario, students either get the answer right or they do not. In this case, there are more answers available to the participants. The examiners were first presented with the latent print, and were asked to classify it into one of three categories: Value for identification, value for exclusion only or of no value. If the examiner found the latent to be of value (for identification or for exclusion only), the examiner was then presented with an exemplar. The decisions available to the examiner at that point were: individualization, exclusion or inconclusive. Thus the analysis is considerably more complicated than that of the hypothetical mathematics exam described above.

They found six false individualizations (0.1% of those regarded as of value for identification). Additionally they found 450 false exclusions (7.5% of those regarded as of value for such evaluation). These results are "absolute results," with the qualifications discussed above. Individuals varied widely in the number of comparisons they actually made, ranging from 13% to 73% among all presentations, 19% to 94% when pairs judged by the examiner to be of no value are excluded and 22% to 100% when limited to pairs regarded as of value for identification (ibid., Table S7). While some of this variability might be explained as the result of the fact that each examiner was presented with a random sample of pairs, it still suggests radical variation in the willingness of participants in the study to commit. (These are "relative results," comparing test-takers to one another.) Other relative results are that uncertified examiners are as accurate as certified examiners [see also Pacheo, Cerchai and Stoiloff (2014)], and that length of experience is uncorrelated with accuracy [see also Thompson, Tangen and McCarthy (2014)]. (The latter is not as strange as it may seem. Examiners rarely get feedback from ground truth, and hence have no opportunity to improve.) Relative results in other studies notably show that training does improve performance [Langenburg, Champod and Genessay (2012), Tangen, Thompson and McCarthy (2011) and Thompson, Tangen and McCarthy (2013, 2014)].

One unfortunate aspect of proficiency testing as it is currently conducted is that the outcome data is sparse—one knows only the conclusion of the analyst, and nothing of the thinking behind it. This issue has been confronted before in other contexts. For example, in an effort to improve the teaching of physics, Larkin et al. (1980) used think-aloud protocols to understand the differences in approach between novice and expert physics problem-solvers. See also Ericsson and Simon (1980, 1993). The closest work of this type I have found related to fingerprint analysis is a paper by Mustonen, Hakkarainen and Tuuainen (2015) in which they report excerpts from the statements of fingerprint analysts who made different conclusions from the same fingerprint comparisons. There are also some annotated comparisons in Neumann et al. (2013). These illuminate some of the different ways experienced analysts go about their work.

g. DNA envy

Until the advent and acceptance of DNA evidence in court, fingerprints were regarded as the most solid of the forensic sciences. In what way, one might ask, are inferences from DNA better grounded than are inferences from fingerprints? In the case of DNA, there is well-established genetic theory, beginning with Mendel, to support the assumption of independence between alleles from different chromosomes, or from alleles on distant parts of the same chromosomes at the individual level. At the population level, it is possible that subpopulations might have positively correlated allele frequencies. However, research has shown that genetic mixing among populations have led to such small estimates of this correlation that it can be neglected for practical purposes [see, for example, Roeder et al. (1998)]. As a result, it is scientifically legitimate to multiply the frequencies of characteristics from different chromosomes. This leads to well-based numerical estimates of how frequently another person would be found with the specified genetic characteristics. It also leads to remarkably small numbers for those estimates when many characteristics are measured.

In the case of fingerprints, the underlying science is absent. There is no particular reason to believe that the finding of a particular characteristic, such as the ending of a ridge in a particular location, is uncorrelated with another characteristic, such as the finding of a bifurcation at some other location. Thus it is not legitimate to multiply the frequencies together. It is possible that a standard way of classifying the Type 1 and 2 characteristics of fingerprints could be found. Applied to a very large database such as the US national database, frequencies of different fingerprint characteristics and their correlations could be estimated. To do this would require serious resources and access to databases not publicly available. Hence fingerprints, with our current scientific knowledge, cannot deliver the kinds of conclusions that DNA can.

h. A digital future?

There is discomfort in some quarters with the thought that fingerprint analysis is subjective. [See President's Council of Advsors on Science and Technology (2016).] Of course it is, and there's nothing intrinsically wrong with that. A fingerprint analyst, accepted in court as an expert, is permitted to give an opinion, which is naturally subjective. However, there is no particular reason why anybody else's subjective opinion should be the same, or even be influenced by the analyst's opinion. The key question is not subjectivity, but rather what evidence can be adduced to support the subjective opinion offered. Thus an opinion is just the opening summary of the real substance, which is the rationale offered for the opinion. A consumer of the opinion, such as a juror or judge, may or may not find the rationale convincing, and hence may or may not accept the proffered opinion as influential to their own.

One way of reducing the worry about subjectivity might be to substitute computers in the place of analysts. If such a computer system identifies the minutiae and other characteristics to be used without human intervention, that would make the inferences more reliable, in the sense that presumably a user would get the same answer from the same inputs each time. However the algorithm embedded in such a computer system would have to deal with the issue of the lack of support for the assumption of independence. The danger in looking to computers for a solution to the issue of subjectivity is that for most lay people, computer code is difficult to understand. Therefore the assumptions underlying the computations would be hidden, but still present. This would be a step away from full disclosure of the basis for a conclusion about a fingerprint question. "The computer says so" ought not to be regarded as a convincing argument.

i. Reporting

There are many varieties of conclusion that have been proposed and used over the years, including:

(a) as an objective fact, the flesh that made this mark at the crime scene is the same flesh that made this print, to the exclusion of all others,
(b) same statement, deleting the last six words,
(c) same as (b), but substitute "It is my opinion that ..." for "As an objective fact, ...",
(d) same as (c) adding "The likelihood the impression was made by another (different) source is so remote that it is considered a practical impossibility" [Scientific Working Group on Friction Ridge Analysis (2011)].

Cole (2014) reviews the history behind these various formulations, and the ambiguities of language underlying them. In 2018, the Department of Justice stated that its latent print examiners may find source identification the conclusion "that the observed friction ridge skin features are in sufficient correspondence such that the examiner would not expect to see the same arrangement repeated in an impression that came from a different source." However "An examiner shall not assert a 100% level of certainty" or "assert that latent print examination is infallible" [Department of Justice (2018)].

Similarly, the Defense Forensic Science Center recently amended the way they report fingerprint analysis results [Swofford (2015)]. The new language is "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." Once again, such a statement, although vague ("is considered extremely low" by whom?), invites the reader or listener to make the unwarranted conclusion that (with high probability) only one person could have made both.

However, this conclusion is unwarranted scientifically, for the reasons given below.

Now I will suppose that the analyst is working under ideal conditions, which include a case manager who makes available to the analyst only the information vital to the comparison of fingerprints (such as the surface from which the mark was lifted). The case manager does not make available any other information about the crime, in order to shield the analyst from biasing information. [Perhaps contextual information could be released sequentially [Dror et al. (2015)].]

I also suppose that the analyst examines first the mark at the crime scene, and records each of the minutiae and other information the analyst intends to compare to the prints. Suppose there is a single print at issue. If there are discrepancies, the analyst must decide whether to declare "exclusion." At one time, a single discrepancy would have been sufficient to occasion such a declaration, but further thought has softened that stance [for discussion, see The Detail (2006)].

Suppose that no discrepancies have been found, and certain coincidences have been found between the mark and the print. The issue is what to make of this. Certainly, less frequent coincidences carry more information, intuitively, than more frequent ones. But unlike the analysis of DNA, the assumption of independence between coincidences is not warranted for fingerprints. Consequently, it would be necessary to know the probability of the joint occurrence of the specific coincidences found. Such data do not exist. At this point the current resources of science are exhausted.

Recent work [Kadane and Koehler (2018)] shows that lay people respond to summaries of the data worded as (a) "identifying the source," (b) "identifying the source to a reasonable degree of scientific certainty" (c) "effect an individualization" and (d) "effect an individualization to the exclusion of all other possible sources in the world" in the same way as stating that this is the person who left the mark. The proposed Defense Department language "the likelihood of observing... is considered extremely low," was somewhat less incriminating. The language "cannot exclude" was least incriminating, and only this language comports with what is scientifically defensible.

6. Conclusion. A fingerprint analyst's finding of minutiae and other indicators common to a mark and a specified source reduces the group of possible creators of the mark. However, the available science does not permit one to say how large the resulting group is. Furthermore, there are no scientifically established criteria for when that set is reduced to a single person. Under these circumstances, I believe that fingerprint reports and testimony should be limited to exclusion (the person whose print is under discussion is very unlikely to have left the mark), inclusion (the person under discussion is one of the people who could have left the mark) and undetermined.

When the comparison of a mark and a print reveal similarities and no differences in minutiae and Level 1 characteristics, I recommend that the analyst report and testify that the defendant is one of the people who could have left the mark at the crime scene. Another way to express this is that the defendant is not excluded. In addition to being scientifically defensible, such testimony obviates discussion of whether fingerprints are unique.

A referee argues for an analogy between fingerprint analysis and eyewitness testimony. When eyewitness testimony concerns identifying someone the testifier knows well, it is certainly due deference. However, eyewitness testimony concerning strangers has met a barrage of criticism [see, for example, Arkowitz and Lillefield (2010)]. There are all sorts of tricks the mind can play on someone trying to reconstruct a memory. Indeed, although I think fingerprint analysts exaggerate the import of their findings, my hunch is that stranger eyewitness testimony is even less reliable.

I do not intend to give the impression that I think fingerprint analysis is worthless. It is clearly useful in eliminating possible suspects. It can also be useful as an investigatory tool to suggest that the perpetrators of several crimes may be the same person. It is also, I believe, relevant at trial as a piece of circumstantial evidence. The current testimonial practices, even those so recently suggested by the Department of Justice, exaggerate the import of fingerprint evidence in a way that will bring it to disrepute.

Acknowledgments. The material presented here is based upon work supported in part under Award No. 70NANB15H176 from the U.S. Department of Commerce, National Institute of Science and Technology. Any opinions, findings or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Institute of Science and Technology, nor the Center for Statistics and Applications in Forensic Evidence. An earlier version of this paper was given at the Annual Meeting of the American Academy of Forensic Sciences in February, 2017.

REFERENCES

ARKOWITZ, H. and LILLEFIELD, S. D. (2010). Why science tells us not to rely on eyewitness accounts: Eyewitness testimony is fickle. and, all too often, shockingly inaccurate. *Sci. Am.* Available at https://www.scientificamerican.com/article/do-the-eyes-have-it/.

CECIL, M. (2014). *Hoover's FBI and the Fourth Estate: The Campaign to Control the Press and the Bureau's Image*. Univ. Press of Kansas, Lawrence, KS.

COLE, S. A. (2005). More than zero: Accounting for error in latent fingerprint identification. *J. Crim. Law Criminol.* **95** 985–1078.

COLE, S. A. (2014). Individualization is dead, long live individualization! Reforms in reporting practices for fingerprint analysis in the United States. *Law Prob. Risk* **13** 117–150.

COMMITTEE ON ASSESSING THE NEEDS OF THE FORENSIC SCIENCE COMMUNITY (2009). *Strengthening the Forensic Sciences in the United States: A Path Forward*. The National Academies Press, Washington, DC.

COMMITTEE ON SCIENTIFIC ASSESSMANT OF BULLET LEAD ELEMENTAL COMPOSITION COMPARISON (2004). *Forensic Analysis Weighing Bullet Lead Evidence*. The National Academies Press, Washington, DC.

COMMITTEE TO REVIEW THE SCIENTIFIC EVIDENCE ON THE POLYGRAPH (2003). *The Polygraph and Lie Detection*. The National Academies Press, Washington, DC.

DEGROOT, M., FIENBERG, S. E. and KADANE, J. B. (1983). *Statistics and the Law*. Wiley, New York.

DEPARTMENT OF JUSTICE (2006). A review of the FBI's handling of the Brandon Mayfield case. Unclassified executive summary. Available at https://oig.justice.gov/special/s0601/final.pdf.

DEPARTMENT OF JUSTICE (2018). Approved uniform language for testimony and reports for the forensic latent print discipline. Available at https://www.justice.gov/file/1037171/download.

DOYLE, C. (1903). The adventure of the Norwood builder. *Strand Magazine* and *Collier's Magazine*, later published as Chapter 2 in "The Return of Sherlock Holmes."

DROR, I. E. and CHARLTON, D. (2006). Why experts make errors. *J. Forensic Identif.* **56** 600–616.

DROR, I. E., CHARLTON, D. and PERON, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci. Int.* **156** 74–78.

DROR, I. E. and ROSENTHAL, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts. *J. Forensic Sci.* **53** 900–903.

DROR, I. E., CHAMPOD, C., LANGENBURG, G., CHARLTON, D., HUNT, H. and ROSENTHAL, R. (2011). Cognitive issues in fingerprint analysis: Inter- and intra-expert consistency and the effect of a 'target' comparison. *Forensic Sci. Int.* **208** 10–17.

DROR, I. E., THOMPSON, W. C., MEISSNER, C. A., KORNFIELD, I., KRANE, D., SAKS, M. and RISINGER, M. (2015). Lettter to the editor—Context management toolbox: A linear sequential unmaksing approach for minimizing cognitive bias in forensic decision making. *J. Forensic Sci.* **60** 1111–1112.

ERICSSON, K. and SIMON, H. A. (1980). Verbal reports as data. *Psychol. Rev.* **87** 215–251.

ERICSSON, K. and SIMON, H. A. (1993). *Protocol Analysis*, *Verbal Reports as Data*. MIT Press, Cambridge, MA.

FAULD, H. (1880). On the skin-furrows of the hand. *Letter in Nature*, October 8, 1880. Available at http://www.clpex.com/Articles/History/Faulds1880.htm.

FEDERAL BUREAU OF INVESTIGATION (1985). *The Science of Fingerprints: Classification and Uses*. Available at www.gutenberg.org/ebooks/19022.

FIENBERG, S. E., ed. (1989). *The Evolving Role of Statistical Assessments as Evidence in Court*. Springer, Berlin.

FIENBERG, S. E. (2005). Forensic science: The nexus of science and the law. In *Arthur M. Sackler Colloquia of the National Academy of Science*, *November* 16–18. Available at http://www.nasonline.org/programs/sackler-colloquia/completed_colloquia/forensic-science-the-nexus-of-science-and-the-law.html.

FIENBERG, S. E. and STRAF, M. (1982). Statistical assessments as evidence. *J. Roy. Statist. Soc. Ser. A* **145** 410–421.

FLORIDA V. S. HAYES (2015). 15CJ109, August 24, 2015. Tampa, Florida.

GALTON, F. (1888). Letter in *Nature*: Personal identification and description. May 25, 1888.

GALTON, F. (1892). *Finger Prints*. MacMillan, London.

HABER, L. and HABER, R. N. (2008). Scientific validation of fingerprint evidence under Daubert. *Law Prob. Risk* **7** 87–109.

HERSCHEL, W. (1880). Skin furrows of the hand. *Letter in Nature*, November 28, 1880. Available at http://galton.org/fingerprints/herschel-1880-nature-furrows.pdf.

INTERNATIONAL ASSOCIATION FOR IDENTIFICATION (1979). Resolution 1979-7.

INTERNATIONAL ASSOCIATION FOR IDENTIFICATION (1980). Resolution 1980-5.

INTERNATIONAL ASSOCIATION FOR IDENTIFICATION (2007). IAI position concerning latent print identification. November 29, 2007. Available at http://onin.com/fp/IAI_Position_Statement_11-29-07.pdf.

INTERNATIONAL ASSOCIATION FOR IDENTIFICATION (2010). Resolution 2010-18.

KADANE, J. B. and KOEHLER, J. J. (2018). Certainty and uncertainty in reporting fingerprint evidence. *Daedalus* (in press).

LANGENBURG, G. (2011). *The Fingerprint Sourcebook* 14: *Scientific Research Supporting the Foundations of Friction Ridge Examinations*. National Institute of Justice, Washington, DC.

LANGENBURG, G. M. (2012). A critical analysis and study of the ACE-V process. Ph.D. thesis, Univ. Lausanne.

LANGENBURG, G., CHAMPOD, C. and GENESSAY, T. (2012). Informing the judments of fingerprint analysts using quality etrics and statistical assessment tools. *Forensic Sci*. *Int*. **219** 183–198.

LARKIN, J., MCDERMOTT, J., SIMON, D. P. and SIMON, H. A. (1980). Expert and novice performance in solving physics problems. *Science* **208** 1335–1342.

LIU, Y. and SRIHARI, S. N. (2009). A computational discriminability analysis on twin fingerprints. In *Proceedings of the Third International Workshop on Computational Forensics*. Springer, The Hague.

MACEO, A. (2009). Friction ridge skin: Morphogenesis and overview. In *Wiley Encyclopedia of Forensic Science*. DOI:10.1002/97804700061589.fsa358.

MOSES, K. R. (2011). *The Fingerprint Sourcebook* 6: *Automated Fingerprint Identification*. National Institute of Justice, Washington, DC. Available at https://www.ncjrs.gov/pdffiles1/nij/225320.pdf.

MUSTONEN, V., HAKKARAINEN, K. and TUUAINEN, J. (2015). Discrepancies in expert decision-making in forensic fingerprint examination. *Forensic Sci*. *Int*. **254** 215–226.

NATIONAL INSTITUTE OF JUSTICE (2011). *The Fingerprint Sourcebook*. National Institute of Justice, Washington, DC. Available at https://www.ncjrs.gov/pdffiles1/nij/225320.pdf.

NATIONAL RESEARCH COUNCIL (2009). *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, DC.

NEUMANN, C., CHAMPOD, C., YOO, M., GENESSAY, T. and LANGENBERG, G. (2013). Improving the understanding and the reliability of the concept of "sufficiency" in friction ridge examination. Award Number 2010-DN-BX-K267, NIJ. Available at https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=271468.

PACHEO, I., CERCHAI, B. and STOILOFF, S. (2014). Miami–Dade study of the ACE-V process: Accuracy and precision in latent fingerprint examinations. National Institute of Justice Final Report Document #248534, Miami–Dade Police Department.

PRESIDENT'S COUNCIL OF ADVSORS ON SCIENCE AND TECHNOLOGY (2016). Report to the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-comparison Methods.

ROEDER, K., ESCOBAR, M., KADANE, J. B. and BALZAS, I. (1998). Measuring heterogeneity in forensic databases using hierarchical Bayes models. *Biometrika* **85** 269–287.

SCIENTIFIC WORKING GROUP ON FRICTION RIDGE ANALYSIS. Standards for examining friction ridge impressions and resulting conclusions. Version 1 (Latent/Tenprint). Available at http://clpex.com/swgfast/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf.

SPECTER, M. (2002). Do fingerprints lie? The gold standard of forensic evidence is now being challenged. *New Yorker* May 27.

SRIHARI, S. N., SRINIVASAN, H. and FANG, G. (2008). Discriminability of fingerprints of twins. *J. Forensic Identif*. **58** 109–127.

SWOFFORD, H. J. (2015). Use of the term "Identification" in latent print. Information Paper.

TANGEN, J. M., THOMPSON, M. B. and MCCARTHY, D. J. (2011). Identifying fingerprint expertise. *Psychol*. *Sci*. **22** 995–997.

THE DETAIL (2006). Available at http://www.clpex.com/legacy/TheDetail/200-299/TheDetail254.htm. Accessed 28 July 2016.

THOMPSON, M. B., TANGEN, J. M. and MCCARTHY, D. J. (2013). Expertise in fingerprint identification. *J. Forensic Sci*. **58** 1519–1530.

THOMPSON, M. B., TANGEN, J. M. and MCCARTHY, D. J. (2014). Human matching performance of genuine crime scene latent fingerprints. *Law Hum. Behav.* **38** 84–93.

TWAIN, M. (1883). A thumb print and what became of it. Later published as Chapter 31 in "Life on the Mississippi."

TWAIN, M. (1894). *The Tragedy of Pudd'nhead Wilson and the Comedy of Those Extraordinary Twins*. American Publishing Company, Hartford, CT.

ULERY, B. T., HICKLIN, A. R., BUSCAGLIA, J. and ROBERTS, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proc. Natl. Acad. Sci. USA* **108** 7733–7738.

YOON, S. and JAIN, A. K. (2015). Longitudinal study of fingerprint recognition. *Proc. Natl. Acad. Sci. USA* **112** 8555–8560.

ZABELL, S. L. (2005). Fingerprint evidence. *J. Law Policy* **13** 143–179.

UNITED STATES CENSUS BUREAU (2017). World population, more than 7.5 billion on May 25, 2017. Available at https://www.census.gov/popclock/world.

DEPARTMENT OF STATISTICS & DATA SCIENCE
AND CSAFE
CARNEGIE MELLON UNIVERSITY
BAKER HALL 232
PITTSBURGH, PENNSYLVANIA 15232
USA
E-MAIL: kadane@stat.cmu.edu