AFTER UNIQUENESS: THE EVOLUTION OF FORENSIC-SCIENCE OPINIONS

BY WILLIAM C. THOMPSON,
JOELLE VUILLE,
FRANCO TARONI &
ALEX BIEDERMANN

BIG CHANGES ARE OCCURRING IN FORENSIC SCIENCE, PARTICULARLY AMONG EXPERTS WHO COMPARE THE PATTERNS FOUND IN FINGER-PRINTS, FOOTWEAR IMPRESSIONS, TOOLMARKS, HANDWRITING, AND THE LIKE. FORENSIC EXAMINERS ARE REACHING CONCLUSIONS IN NEW WAYS AND CHANGING THE LANGUAGE THEY USE IN REPORTS AND TESTIMONY. THIS ARTICLE EXPLAINS THESE CHANGES AND THE CHALLENGES THEY POSE FOR LAWYERS AND JUDGES.

Although testimony about forensic comparisons has been offered in court for over a century, it has recently become controversial. Questions have emerged about the scientific foundation of the pattern-matching disciplines and about the logic underlying forensic scientists' conclusions. The traditional assumption that items like fingerprints and toolmarks have unique patterns that allow experts to accurately determine their source has been challenged and is being replaced by a new logic of forensic reporting. The new logic requires experts to evaluate and weigh probabilities rather than claim certainty. Forensic experts must now moderate the claims they make about their own accuracy and, increasingly, use numbers to describe the strength of their conclusions. Because these changes have important implications for the probative value of the conclusions that forensic experts offer in court, it is important that judges understand them.

THE DEMISE OF THE THEORY OF DISCERNIBLE UNIQUENESS

As recently as a decade ago, forensic scientists in the pattern-matching disciplines told a common story when asked to explain how they reached conclusions. Their analytic process began with the assumption that the items they examined had unique patterns: For example, every finger was said to have a unique set of friction ridges, and thus every print left by a given finger (if sufficient in size and clarity) was expected to be different from the print made by any other finger. Similarly, every gun barrel was thought to be unique; hence the pattern of marks found on bullets fired through a given barrel (if sufficient in size and clarity) was expected to differ from the pattern found on bullets fired through any other gun barrel. The soles of shoes and human dentition also were presumed to be unique, and thus the impressions left by a given shoe, or a given set of teeth (if sufficiently clear and detailed) were assumed to differ from the impressions left by any other shoe or set of teeth. Applying the same analysis, everyone's handwriting was presumed to be unique, and hence a sample of handwriting from a given individual (if sufficiently extensive) was presumed to be distinguishable from the handwriting of any other individual. These presumptions have been called *the theory of discernible uniqueness*.[1]

According to this traditional account, the job of the forensic examiner was first to assess whether the patterns seen in impressions contained sufficient detail to allow a determination of source and, second, to compare the impression patterns. If sufficient detail was available, then a "match" between the patterns meant the source of the impressions must necessarily be the same, and a mismatch (failure to match) meant that the source of the impressions must necessarily be different. If insufficient detail was available to make a definitive determination, then the examination was inconclusive.

Examiners in a number of forensic disciplines have testified that this analysis allows them to make source determinations with complete certainty. A prominent fingerprint examiner explained the matter as follows:

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.[2]

Unfortunately, these claims have not withstood scientific scrutiny. Indeed, commentary on the issue in the broader scientific and academic communities (beyond the community of forensic science practitioners) has been nearly unanimous in dismissing such claims as unwarranted.[3]

Consider the claim that the ridge patterns on every finger are unique. Like similar claims about snowflakes, it is impossible to demonstrate empirically that this claim is true because one cannot conduct a systematic comparison of every finger against every other. Furthermore, there is a difference between the claim that the ridge pattern on each finger is unique and the claim that a fingerprint examiner can accurately determine whether two fingerprints were made by the same finger. The validity of the latter also depends on the quality of the prints and the level of analysis employed during the comparison. Even if the ridge detail of every finger were unique, it does not follow that every impression made by every finger will always be distinguishable from every impression made by any other finger, particularly when the impressions are of poor quality (e.g., limited detail, smudged, distorted, or overlaid on another impression). By analogy, it may be that every human face is unique, but we can still mistake one person for another, particularly when comparing poor-quality photos.[4]

This is a limitation that most fingerprint examiners now acknowledge:

When fingerprint comparisons are being made, they are not being made from friction ridge skin to friction ridge skin. They 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.[5]

Consequently, the key scientific question is not whether the ridge pattern of ▶

each finger is unique, but how well an examiner can distinguish the impressions of different fingers at the level of analysis applied in a forensic examination. That question cannot be answered by assertions about the uniqueness of ridge patterns; it can only be answered by empirical research.

This critique also applies to other forensic pattern-matching disciplines, such as toolmark analysis, footwear analysis, handwriting analysis, and bitemark analysis. Although some practitioners in these fields persist in making the injudicious claim that their conclusions *must* be accurate because they are comparing patterns that are unique, the broader scientific community has called for empirical studies to put such claims to the test.

A key event in the evolution of forensic science opinion was a 2009 report by the United States National Academy of Sciences (NAS), which called for the development of "quantifiable measures of the reliability and accuracy of forensic analyses" that reflect "actual practice on realistic case scenarios . . . ."[6] It called for research to establish "the limits of reliability and accuracy that analytic methods can be expected to achieve as the conditions of forensic evidence vary."[7] The report concluded that "much forensic evidence — including, for example, bitemarks and firearm and tool mark identifications — is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline."[8]

In response to this high-level scientific criticism, forensic scientists made some efforts to study the accuracy of their methods, although these efforts have been limited. The FBI commissioned an important series of studies on the accuracy of latent print anal-

ysis, but relatively little research has been conducted on the accuracy of other forensic science disciplines. In 2016, the President's Council of Advisors on Science and Technology (PCAST) issued a report that reviewed scientific research published to that point on the accuracy of six forensic science disciplines that rely on "feature comparison": DNA analysis, latent print analysis, firearms analysis, bitemarks analysis, footwear analysis, and microscopic hair analysis.[9] PCAST found that adequate research had been done to establish the "foundational validity" of latent print analysis and DNA analysis of single-source and simple mixture samples. "Foundational validity" means the method in question is capable of producing accurate results when properly performed. PCAST concluded, however, that too little research had been published to establish the "foundational validity" of firearms analysis, bitemarks analysis, footwear analysis, microscopic hair analysis, and DNA analysis of complex mixtures.

Moreover, even if latent print examination has "foundational validity," the studies do not show that it is infallible (as examiners have claimed). The studies reviewed by PCAST showed that latent print examiners have:

> . . . a false-positive rate that is substantial and is likely to be higher than expected by many jurors based on longstanding claims about the infallibility of fingerprint analysis. The false-positive rate could be as high as {one} error in 306 cases {based on an FBI study} and {one} error in 18 cases based on a study by another crime laboratory.[10]

The studies reviewed by PCAST also showed substantial numbers of false exclusions.[11]

In light of these developments, forensic scientists have begun to change the way they describe their analytic process and report their conclusions. They can

no longer credibly claim the ability to infallibly discern whether two compared sets of features share a unique pattern and thus have a common source. Professional associations and standards-setting bodies in various branches of forensic science have recommended that examiners avoid asserting that their conclusions are infallible and avoid claiming that they can discern whether a pattern is unique.[12] Experts are now discussing a variety of new approaches to reporting.

THE LOGIC OF FORENSIC INFERENCE

To understand and evaluate the new approaches to reporting, it is necessary to understand the logic of forensic inference — that is, the logical steps by which a forensic examiner proceeds from observations to conclusions. Let's consider, as an example, the logical steps that lead a latent print examiner from the observation that two fingerprints have similar ridge patterns to conclusions about whether they were made by the same finger. If examiners can no longer credibly claim that prints must necessarily have a common source if they appear to have "matching" ridge patterns, what conclusions can they reasonably draw?

The new approaches all recognize that forensic inference requires an inductive line of reasoning, which entails consideration of probabilities. The examiner must consider the probability of seeing the patterns observed in the impressions under two alternative hypotheses about their origin: (1) that the impressions have the same source (e.g., same finger, same tool); and (2) that the impressions have a different source.

Suppose, for example, that a latent print examiner observes that two fingerprints have similar patterns but with slight discrepancies. The examiner must consider how probable it would be to observe those particular patterns

(including both similarities and discrepancies) *if* the prints were made by the same finger. This might involve consideration of the likelihood that slipping or torsion of the finger, or some other process, could have distorted one or both of the prints enough to produce the discrepancies. The examiner must also consider how probable it would be to observe those particular patterns (including both similarities and discrepancies) *if* the prints were made by different fingers. This would involve consideration of the rarity of the shared features, hence how likely or unlikely it would be to observe so much similarity in prints made by different fingers.

In order to draw inferences and reach conclusions about whether two impressions have a common source, the expert must consider the balance between the two key probabilities: (1) the probability of the observed patterns if the impressions have the *same* source; and (2) the probability of the observed patterns if the impressions have a *different* source. The ratio between these two probabilities provides an index of the probative value of the evidence for distinguishing the two hypotheses. The evidence favors a particular hypothesis to the extent that the observed results are more probable under that hypothesis than under the alternative hypothesis. For example, if a latent print examiner thinks the observed ridge patterns (including both similarities and discrepancies) would be more probable if the prints have the same source (same finger) than if they have a different source (different fingers), then the evidence supports the hypothesis that the prints have the same source.

This logic is fundamental and inescapable. It is the basis for any conclusions that examiners choose to report.

APPROACHES TO REPORTING

There are several schools of thought

PROFESSIONAL ASSOCIATIONS AND STANDARDS-SETTING BODIES IN VARIOUS BRANCHES OF FORENSIC SCIENCE HAVE RECOMMENDED THAT EXAMINERS AVOID ASSERTING THAT THEIR CONCLUSIONS ARE INFALLIBLE AND AVOID CLAIMING THAT THEY CAN DISCERN WHETHER A PATTERN IS UNIQUE.

about how examiners should report their conclusions regarding the balance of probability. In this section of the article, we will outline the different approaches and discuss their strengths and weaknesses.

Likelihood Ratios. One approach that is popular in Europe allows examiners to use numbers called likelihood ratios to describe their perception of the balance of probabilities.[13] The likelihood ratio represents the expert's view of the relative probability of the observed features under the alternative hypotheses about the source of the impressions. A likelihood ratio of 1000, for example, represents the expert's view that the observed patterns are 1000 times more probable under one hypothesis (e.g., same source) than under the alternative hypothesis. Experts typically make the favored hypothesis the numerator of the likelihood ratio so that reported values range from one to infinity. A value of one

means the expert thinks the observed patterns are equally likely under the two hypotheses, and hence the evidence has no value for distinguishing the hypotheses. A value greater than one means the expert thinks the observed patterns are more likely under one hypothesis than the alternative, and thus the forensic evidence supports the favored hypothesis. The larger the likelihood ratio, the greater the expert's perception of how strongly the balance of probabilities supports the favored hypotheses. European latent print experts sometimes report very high likelihood ratio values, such as one million or even ten million.

The European Network of Forensic Science Institutes (ENFSI) and the U.K. Royal Statistical Society promote the use of likelihood ratios to describe experts' assessments of the strength of forensic evidence.[14] Many forensic scientists in Europe, New Zealand, and parts of Australia also have adopted this approach.[15]

The question most commonly asked about likelihood ratios is how the experts come up with the numbers they report. In some disciplines, experts can rely on databases and statistical modeling. This is most common in fields like forensic DNA analysis and forensic voice comparison, where extensive databases exist and methods for statistical modeling have been evaluated in the scientific literature.[16] Likelihood ratios have been presented in the United States for many years in connection with forensic DNA evidence. The expert typically says something like the following:

The genetic characteristics observed in the evidentiary sample are X times more likely if the defendant was a contributor than if the contributor was instead a random unknown Caucasian.

In the past, there has been insufficient data on the rarity of the features observed by experts in most pattern-matching ▸

disciplines to allow statistical estimates, but that is starting to change. Recently the Defense Forensic Science Center (DFSC) of the Department of the Army began presenting probabilities in connection with fingerprint evidence. In March 2017, the laboratory announced that future reports would include statements like the following:

> The latent print on Exhibit ## and the standards bearing the name XXXX have corresponding ridge detail. The probability of observing this amount of correspondence is approximately ## times greater when impressions are made by the same source rather than by different sources.[17]

The laboratory uses a software program to score the similarity of the prints being compared based on "the spatial relationship and angles of the ridge details."[18] The program then uses a database to evaluate how much more common it is to observe a given similarity score when comparing prints from the same finger than prints from different fingers. Although this is a novel method that has not yet been adopted by other forensic laboratories, the DFSC has reportedly offered to share this software with any government forensic laboratory in the United States, and other labs are evaluating this approach. Similar software-based, quantitative methods for assessing toolmark and handwriting evidence also are under development, although it may be a few years before they are ready for the courtroom. As experts begin offering testimony based on these new methods in United States courtrooms, litigants are likely to challenge admissibility under the *Daubert* or *Frye* standards, which will require judges to scrutinize whether the new methods are reliable and generally accepted.

Likelihood ratios also can be reported in forensic science disciplines that have

IF THE EXAMINER DOES NOT KNOW ENOUGH TO ASSESS THE RELEVANT PROBABILITIES, THEN THE EXAMINER DOES NOT KNOW ENOUGH TO EVALUATE THE STRENGTH OF THE FORENSIC EVIDENCE — AND HENCE NOTHING THE EXAMINER SAYS ABOUT THE VALUE OF THE EVIDENCE SHOULD BE TRUSTED.

not developed databases and statistical models. In those fields, experts may rely on their training and experience to come up with a likelihood ratio. In some instances, a likelihood ratio can be based partly on empirical data and partly on the expert's judgment.[19] While some commentators have derided such estimates as "subjective" and questioned their validity (one commentator called them "numbers from nowhere"[20]), the practice of presenting likelihood ratios based on expert judgment (rather than a database) appears to have taken hold in many European countries.[21] Whether such testimony should be admitted in the United States is an issue judges may soon need to contemplate.

Those who support the use of likelihood ratios based on expert judgment (rather than databases) point out that a forensic examiner must make subjective judgments of probability in order to draw *any* conclusions about whether

two items have a common source.[22] If the examiner does not know enough to assess the relevant probabilities, then the examiner does not know enough to evaluate the strength of the forensic evidence — and hence nothing the examiner says about the value of the evidence should be trusted. It makes no sense, proponents say, to allow experts to testify about conclusions they reached based on a subjective judgment of the balance of probabilities but not allow the expert to use a likelihood ratio to say what their judgment was. When experts report their judgments of the likelihood ratio, proponents argue, the expert's judgmental process is more transparent, and hence the value of the expert's conclusions is easier to evaluate.[23]

Verbal Equivalents of Likelihood Ratios. Examiners may nevertheless be reluctant to put specific numbers on their subjective judgments, even if those judgments are well grounded. An examiner may justifiably believe that the observed results are more probable if the items being compared have the same source than a different source, for example, without being able to say with any precision how much more probable. Forcing examiners to articulate numbers may lend a false air of precision to a subjective assessment.

One way to avoid this problem is to allow examiners to express conclusions about the balance of probabilities using words rather than numbers. In a 2012 report, a group of experts assembled by the National Institute of Standards and Technology (NIST) recommended that latent print examiners report their conclusions using statements like the following:

> It is far more probable that this degree of similarity would occur when comparing the latent print with the defendant's fingers than with someone else's fingers.[24]

TABLE 1. PROPOSED LIKELIHOOD RATIO TERMINOLOGY *(AFSP, 2009)*	
NUMERICAL EXPRESSION OF PROBATIVE STRENGTH *(likelihood ratio)*	VERBAL EXPRESSION OF PROBATIVE STRENGTH
1 – 10	Weak or limited
10 – 100	Moderate
100 – 1,000	Moderately strong
1000 – 10,000	Strong
10,000 – 1,000,000	Very strong
> 1,000,000	Extremely strong

This approach allows examiners to substitute an imprecise verbal statement ("far more probable") for a number, while still explaining the strength of the forensic evidence in terms of the balance of probabilities. Of course lawyers can (and should) ask experts testifying in this manner to explain what they mean by statements like "far more probable" and what basis they have for that conclusion.

Another approach that has been popular in Europe substitutes words for numerical likelihood ratios. The U.K.-based Association of Forensic Science Providers (AFSP) has proposed that forensic scientists use the "verbal expressions" shown in Table 1 (above) to describe how strongly their evidence supports a particular hypothesis about the evidence (e.g., the hypothesis that two items have a common source).[25] Under this approach, forensic scientists first come up with a likelihood ratio that reflects their perception of the balance of probabilities, and then use one of the verbal expressions in the table instead of (or in addition to) the number to describe their conclusions in reports and testimony.

For example, a forensic scientist who concludes (by whatever means) that the results observed in a forensic comparison are 500 times more likely if the items have a common source than if they have a different source would report that the comparison provides "moderately strong" support for the conclusion that the items have a common source. A forensic scientist who concluded that the results are 100,000 times more likely if the patterns being compared have a common source would say that the evidence provides "very strong support" for the hypothesis of a common source. Statements of this type are not common in U.S. courts, but they have been discussed extensively in the academic literature.[26] They offer one possible answer to the question of how to report source conclusions.

Match Frequencies / Random Match Probabilities. When a comparison reveals matching features in two items, forensic scientists sometimes estimate and report the frequency of the matching features in a reference population. This occurs most commonly in forensic DNA analysis, where genetic databases provide an empirical basis for assessing the proportion of a population that has a particular genetic feature. Forensic DNA analysts sometimes refer to these estimates as *match frequencies* (e.g., "The blood stain at the crime scene and the reference blood sample from the suspect have the same DNA profile. This profile is estimated to occur in one person in 10 million among Caucasian-Americans."). Alternatively, they may present these estimates as *random match probabilities (RMPs)* (e.g., "The probability that a random Caucasian-American would match this DNA profile is 0.0000001 or 1 in 10 million."). As forensic scientists develop databases that can be used to quantify the rarity of pattern features,

we are likely to see similar testimony in other pattern-matching disciplines.

Even without empirical data, experts sometimes make statements about the random match probability based on training and experience. These subjective-match probabilities are typically reported with words rather than numbers. An examiner might say, for example, that the set of features shared by two items is "rare" or "unusual."

One drawback of this approach is that it addresses only one of the two questions needed to evaluate the balance of probabilities reflected in the likelihood ratio. It addresses the probability of the observed patterns under the hypothesis that they have a different source. It fails to consider the probability of the observed patterns if the impressions have the same source. Consequently, this approach may be misleading in cases in which the latter probability is low, when, for instance, the patterns have important discrepancies as well as similarities. Likelihood ratios, which consider both probabilities, arguably offer a more balanced and complete account of the strength of such evidence.

Source Probabilities. In the United States, forensic examiners often present opinions on the probability that two items have a common source. Opinions of this type can be expressed quantitatively, using probabilities or percentages. For example, a forensic scientist might say there is a 99 percent chance that two items have a common source. It is more common, however, for examiners to express such conclusions with words rather than numbers. For example, the forensic scientist might say it is "moderately probable," "highly probable," or "practically certain" that two items have a common source.

Lawyers and judges tend to like source probabilities because they are ▶

easy to understand; they address the exact question that the trier of fact needs to assess — how likely it is that the two impressions (e.g., two finger-prints) come from the same source? The problem, unfortunately, is that the information forensic scientists can glean from a comparison of impressions is not, by itself, sufficient to allow them to reach conclusions about source prob-ability. As we will explain, examiners can logically draw conclusions about source probabilities *only* by combining conclusions drawn from a comparison of the impressions with assumptions or conclusions about the strength of other evidence that bears on the ques-tion of whether the impressions being compared have a common source.[27]

To illustrate, consider the Elvis Problem discussed in the sidebar. What is the probability that Elvis Presley was the source of the evidence left at the crime scene? As explained, this ques-tion cannot be answered based on the forensic science evidence alone. It is only by making assumptions or drawing conclusions about the likelihood of Elvis being at the crime scene — a matter having nothing to do with the foren-sic science evidence — that the forensic examiner can draw conclusions about the probability that Elvis was the source.

The same problem arises whenever forensic scientists express opinions on source probabilities. The opinion must, of logical necessity, depend in part on conclusions or assumptions about matters having nothing to do with forensic science, such as whether the person who is alleged to have left a trace (e.g., a fingerprint or shoeprint) at the crime scene is a likely or unlikely suspect and how many other people had access to the crime scene. Forensic examiners are not in a good position to make such judgments and have no busi-ness doing so anyway.

IN ORDER TO UNDER-STAND THE EXPERT'S CONCLUSIONS, THE TRIER-OF-FACT WILL NEED TO KNOW HOW THE EXPERT EVALUATED THE RELEVANT PROBABILITIES, AND HOW, WHERE, AND WHY THE EXPERT SET THE THRESHOLD FOR REPORTING A PARTICULAR DECISION. THE TRIER-OF-FACT ALSO WILL NEED INFORMATION ABOUT THE ACCURACY OF DECISIONS REACHED IN THIS MANNER.

Identification and Exclusion. In the United States, the most popular method of reporting results of forensic compari-sons is to state a bottom-line conclusion about whether two traces have a common source. The conclusion that two traces have the same source is often described as "identification" or "individualiza-tion," while a conclusion that they have a different source is "exclusion." These conclusions can be seen as extreme examples of source probabilities, corre-sponding to either a 100 percent or a zero percent chance that the traces being compared have the same source.

The demise of the theory of discernible uniqueness has made these conclusions more difficult to justify. Most experts now acknowledge that these conclusions require the examiner to *make a decision* about whether the evidence is strong enough to support a definitive conclu-sion, but there does not appear to be a generally accepted theory regarding how experts should make that decision.

One approach requires experts to make an assessment of the source proba-bility. They report "identification" when their assessed source probability exceeds some high threshold and "exclusion" when their assessment falls below some low threshold. As discussed in the previ-ous section, however, the assessment of source probabilities requires the expert to make assumptions or draw conclu-sions about matters beyond the forensic comparison in question. Experts cannot draw conclusions about source probabil-ities without facing the Elvis Problem, which renders such conclusions prob-lematic. If courts allow experts to present conclusions reached in this manner, they should also require experts to disclose the factual basis for their asserted source probabilities. To evaluate the expert's conclusion, the trier-of-fact will need to know the extent to which the expert's decision was influenced by assumptions or conclusions about matters beyond the realm of forensic science.

To avoid the Elvis Problem, foren-sic scientists might instead base their decision on their judgment of the balance of probabilities. If they believe the balance weighs strongly enough in favor of the hypothesis that the items being compared have the same source, then they might report "identifica-tion." If they believe the balance weighs strongly enough in favor of the hypoth-esis that the items have a different source, then they might report "exclu-sion." This approach avoids the need for ▸

ELVIS'S ALIBI

IMAGINE THAT A BLOODSTAIN OF RECENT ORIGIN IS FOUND AT THE SCENE OF A CRIME. IMAGINE FURTHER THAT THE DNA PROFILE OF THE BLOODSTAIN IS SOMEHOW DETERMINED TO BE THE SAME AS THE DNA PROFILE OF ROCK-AND-ROLL LEGEND ELVIS PRESLEY. FINALLY, IMAGINE THAT THE DNA PROFILE IN QUESTION IS ONE MILLION TIMES MORE LIKELY TO BE OBSERVED IF THE SAMPLE CAME FROM ELVIS THAN IF IT CAME FROM A RANDOM PERSON. BASED ON THE DNA EVIDENCE, WHAT CAN THE EXAMINER LOGICALLY INFER ABOUT THE PROBABILITY THAT THE CRIME SCENE STAIN CAME FROM ELVIS PRESLEY?

A moment of reflection should be sufficient to realize that the examiner can draw no conclusion about the probability that the crime scene stain came from Elvis based on the DNA evidence alone; the examiner must also consider other matters, such as whether Elvis could plausibly be the source. In this case, the suspect (Elvis) has a strong alibi – he was widely reported to have died in 1977. If the forensic scientist believes this "alibi," then the probability that the bloodstain came from Elvis is necessarily zero.

An examiner who believes Elvis is dead might decide to report that there is a zero percent chance the crime scene sample came from Elvis. Notice, however, that this conclusion is not based on the strength of the DNA evidence. It depends entirely on the expert's assessment of matters beyond the realm of forensic science – in this case Elvis's alibi.

The expert might try to take a neutral position on the alibi – assuming, for example, that the question of whether Elvis could have been the source is a toss-up or 50:50 chance. When this seemingly neutral assumption about the truth of the alibi is taken as a starting point, the expert can update the initial assessment in light of the DNA evidence. That approach leads logically to the conclusion that there is more than a 99 percent chance that Elvis was the source of the blood.[32] Notice, however, that this conclusion depends only partly on the DNA evidence; it also depends critically on the assumption of a 50 percent chance *a priori* that the blood at the crime scene came from Elvis (an assumption many people will view as fanciful). Should forensic scientists be basing their conclusions on assumptions of this type?

The problem (as should now be clear) is that no assumption about the probability of an alibi's veracity can truly be considered "neutral." Yet without some assumption about the probability of the alibi's veracity, there is no logical way to assess the probability that Elvis was the source.

This same logical conundrum arises in any case in which a forensic scientist is asked to assess the probability that a particular suspect was the source of a crime scene sample based on a forensic comparison. The expert can never answer the question based solely on the forensic evidence. Inevitably the expert must make assumptions or take a position on other matters, such as the probability that the suspect's alibi is true. Doing that may well invade the jury's province; it certainly requires the expert to delve into matters beyond his or her scientific expertise. Consequently, judges should consider carefully whether to admit statements about source probabilities into evidence. If such statements are admitted, judges (and lawyers) should try to make clear to the jury the extent to which the expert's conclusions depend on comparison of the items in question, and the extent to which they depend on assumptions or conclusions about other matters.
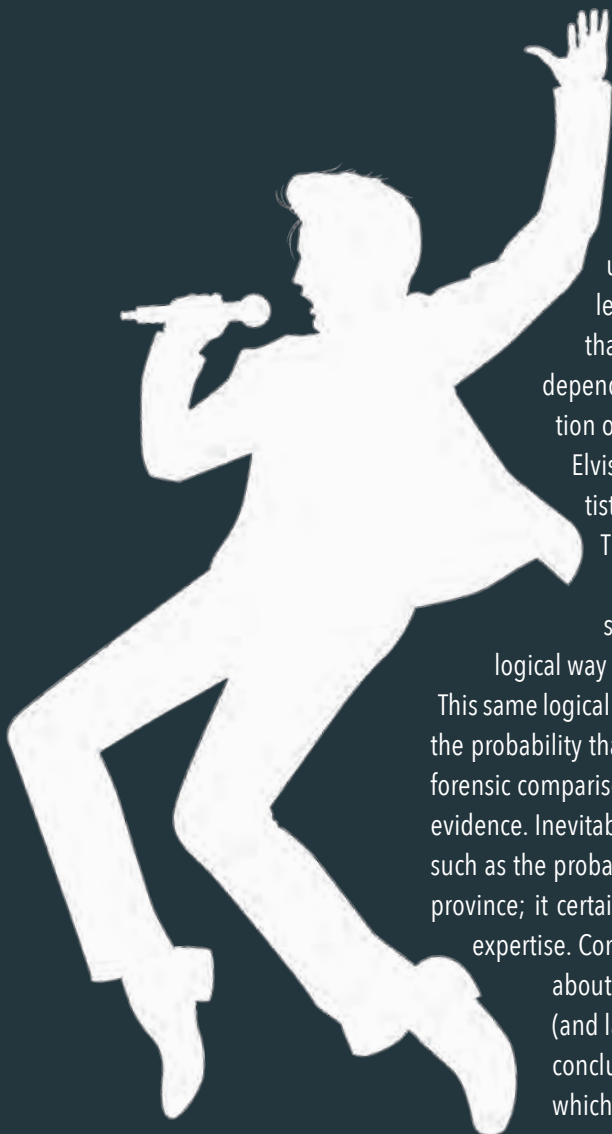
the expert to evaluate source probabilities, but it still raises many questions. In order to understand the expert's conclusions, the trier-of-fact will need to know how the expert evaluated the relevant probabilities, and how, where, and why the expert set the threshold for reporting a particular decision. The trier-of-fact also will need information about the accuracy of decisions reached in this manner.

In the past, expert forensic science testimony about "identification" and "exclusion" often went unchallenged, with lawyers on both sides assuming such testimony was reliable and uncontroversial. As lawyers become more aware of the issues discussed in this article, we expect they will examine the logic and basis of such conclusions far more closely than they have in the past.

THE GROWING IMPORTANCE OF STATISTICAL DATA ON ERROR RATES

Regardless of how forensic scientists choose to present their conclusions, we also expect in the near future to see more testimony about the error rates of pattern-matching disciplines. The 2016 PCAST report argued forcefully that empirical research is the only way to assess the accuracy (and hence the probative value) of examiners' source conclusions:

Without appropriate estimates of accuracy, an examiner's statement that two samples are similar — or even indistinguishable — is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact. Nothing — not training, personal experience nor professional practices — can substitute for adequate empirical demonstration of accuracy.[28]

PCAST called for a continuing program of research in which examiners are tested by having them compare samples from known sources. PCAST recommended that the samples used in the research be representative of the samples encountered in casework, that examiners have no information about the correct answer, that independent groups with no stake in the outcome conduct multiple studies, and that the data be available to other scientists for review.[29] Courts will need to consider the results of such studies when deciding whether testimony about forensic comparisons is sufficiently trustworthy to be admitted — whether, in the words of Rule 702(c) of the Federal Rules of Evidence, it is "the product of reliable principles and methods."[30]

When such testimony is admitted, error-rate data will be relevant for assessing its probative value. PCAST suggested that testimony about error rates of the relevant forensic method, as research has shown on samples like those in the case at hand, should always be presented in conjunction with testimony about the results of forensic comparisons. Experts are likely to be asked about error rates during cross-examination even if the proponent of the forensic evidence elects not to present error-rate data in direct testimony. Lawyers are likely to debate the implications and significance of error-rate data for evaluating the probability that an error occurred in the case at hand.

ALEX BIEDERMANN is associate professor at the University of Lausanne (Switzerland) in the Faculty of Law, Criminal Justice and Public Administration. Previously, he worked as a scientific advisor in the Federal Department of Justice and Police in Berne in cases investigated by the Office of the Attorney General of Switzerland.

WILLIAM C. THOMPSON is professor emeritus of criminology, law and society at the University of California, Irvine. He studies human judgment and decision making with a particular focus on cognitive and contextual bias in scientific assessments, the use of scientific and statistical evidence in the courtroom, and lay perceptions of scientific and statistical evidence.

JOËLLE VUILLE is a senior researcher at the University of Lausanne School of Criminal Justice. She has written extensively on the use of scientific evidence in criminal courts and the use of probabilities in the judicial context.

FRANCO TARONI is professor of forensic statistics at the University of Lausanne, also in the Faculty of Law, Criminal Justice and Public Administration. He has authored and coauthored several books on the use of probabilities in the judicial context and is editor of the journal *Law, Probability and Risk* (Oxford University Press).

We are thus on the cusp of a new era for forensic science — an era in which statistics will inevitably play a greater role. Oliver Wendell Holmes once declared that "the man of the future is the man of statistics . . . . "[31] In the pattern-matching disciplines of forensic science, that future has arrived.

[1] Michael J. Saks & Jonathan Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 Science 892 (2005), at 892.

[2] Heidi Eldridge, *The Shifting Landscape of Latent Print Testimony: An American Perspective*, 3 J. of Forensic Sci. & Med. 72 (2017), at 72.

[3] *See*, Nat'l Acad. of Sci., Nat'l Research Council, Strengthening Forensic Science in the United States: A Path Forward (2009) [hereinafter NAS Report] at 44, 108, 162, 169, 176; President's Council of Advisors on Sci. & Tech., Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods (2016) [hereinafter: PCAST Report], at 19, 30, 54.

[4] Simon Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 Law Probability & Risk 233 (2009), at 236-237.

[5] Eldridge, *supra* note 2, at 76.

[6] NAS Report, *Recommendation 3(b)*, *supra* note 3, at 23.

[7] *Id.*

[8] *Id.*, at 108.

[9] PCAST Report, *supra* note 3.

[10] *Id.*, at 9–10.

[11] *See e.g.*, Igor Pacheco, et al., Miami-Dade research study for the reliability of the ACE-V process: Accuracy & precision in latent fingerprint examinations (2014), at 53-55.

[12] *See e.g.*, Sci. Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), Document # 4: Guideline for the Articulation of the Decision-Making Process for the Individualization in Friction Ridge Examination (Latent/Tenprint) r. 11.2.3 (2013), at 11.2.3.

[13] Colin G.G. Aitken & Franco Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists (2d ed. 2004), at 95 (providing a simple mathematical description of the likelihood ratio that lawyers and judges may encounter when reviewing forensic evidence. Let E represent the observed features of two traces that a forensic scientist is asked to compare; let H_S represent the proposition (hypothesis) that the items have the same source and H_d the proposition that they have a different source. The likelihood ratio is then $p(E|H_S)/p(E|H_d)$, which is read as "the probability of E given H_S over the probability of E given H_d").

[14] European Network of Forensic Sci. Insts., Guideline for Evaluative Reporting in Forensic Science (2015), at 2.4; *see also*, Royal Statistical Soc'y, http://www.rss.org.uk/practitioner-guides (last visited Jan. 7, 2018) (providing reports on this issue).

[15] Alex Biedermann, et al., *Development of European Standards for Evaluative Reporting in Forensic Science: The Gap Between Intentions and Perceptions*, 21 The Int'l J. of Evidence & Proof 14 (2017), at 26.

[16] *See*, John Butler, Fundamentals of Forensic DNA Typing (2009); Geoffrey S. Morrison & William C. Thompson, *Assessing the Admissibility of a New Generation of Forensic Voice Comparison Testimony*, 18 Colum. Sci. & Tech. L. Rev. 326 (2017).

[17] Def. Forensic Sci. Ctr., Dep't of the Army, Information Paper: Modification of Latent Print Technical Reports to Include Statistical Calculations (2017), at 2.

[18] *Id.*, at 2.

[19] Alex Biedermann, et al., *How to Assign a Likelihood Ratio in a Footwear Mark Case: An Analysis and Discussion in the Light of R v T*, 11 Law, Probability & Risk 259 (2012), at 265-270.

[20] D. Michael Risinger, *Reservations About Likelihood Ratios (and Some Other Aspects of Forensic 'Bayesianism')*, 12 Law, Probability & Risk 63, 72 (2012).

[21] Charles E. H. Berger, et al., *Evidence Evaluation: A Response to the Court of Appeal Judgment in R v T*, 51 Sci. & Just. 43 (2011), at 43-44.

[22] Marjan Sjerps & Charles E. Berger, *How Clear is Transparent? Reporting Expert Reasoning in Legal Cases*, 11 Law, Probability & Risk 317 (2012).

[23] *Id.*; Biedermann, *supra* note 19, at 259; William C. Thompson, *Discussion Paper: Hard Cases Make Bad Law — Reactions to R v T*, 11 Law, Probability & Risk 347 (2012), at 351-353.

[24] Expert Working Grp. on Human Factors in Latent Print Analysis, Latent Print Examination and Human Factors: Improving the Practice Through a Systems Approach (2012), at 134.

[25] Ass'n of Forensic Sci. Providers, *Standards for the Formulation of Evaluative Forensic Science Expert Opinion*, 49 Sci. & Just. 161 (2009), at 163.

[26] Raymond Marquis et al., *Discussion on How to Implement a Verbal Scale in a Forensic Laboratory: Benefits, Pitfalls and Suggestions to Avoid Misunderstandings*, 56 Sci. & Just. 364 (2016).

[27] *See*, Bernard Robertson, et al., Interpreting Evidence – Evaluating Forensic Science in the Courtroom (2d ed. 2016), at 16-18; Ian W. Evett, *Towards a Uniform Framework for Reporting Opinions in Forensic Science Casework*, 38 Sci. & Just. 198 (1998), at 200-201 (explaining that after comparing two items, a forensic examiner may be able to judge the probability of the observed results under the alternative hypotheses: $p(E|H_S)$ and $p(E|H_d)$. But these probabilities are not the same as source probabilities; source probabilities are the inverse of these conditionals — i.e., $p(H_S|E)$ and $p(H_d|E)$. To infer source probabilities from the probability of the observed evidence, the examiner must take into account the prior probability that the items have the same source, $p(H_S)$, or different source, $p(H_d)$.).

[28] PCAST Report, *supra* note 3, at 46.

[29] *Id.*, at 66.

[30] Fed. R. Evid. 702(c).

[31] O.W. Holmes, *The Path of the Law, 8* Harv. L. Rev. 457, 469 (1897).

[32] *E.g.*, David J. Balding & Christopher D. Steele, Weight-of-Evidence for Forensic DNA Profile (2015).