Forensic Pattern Recognition Evidence
An Educational Module

Prepared by

Simon A. Cole
Professor, Department of Criminology, Law, and Society
Director, Newkirk Center for Science and Society
University of California, Irvine

Alyse Berthental
PhD Candidate, Department of Criminology, Law, and Society
University of California, Irvine

Jaclyn Seelagy
Scholar, PULSE (Program on Understanding Law, Science, and Evidence)
University of California, Los Angeles School of Law

For

Committee on Preparing the Next Generation of Policy Makers for Science-Based Decisions
Committee on Science, Technology, and Law

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

June 2016

Contents

Introduction

GOALS AND METHODS

The goal of this module is to use forensic pattern recognition evidence as a case study through which to teach professional and other students key concepts in the evaluation of scientific evidence. The popular appeal of forensic evidence makes it well suited for a case study that will engage students. In addition, recent legal and scientific controversies over forensic evidence have generated a wealth of materials that can be used to prompt students to hone their critical faculties in evaluating scientific evidence. The module uses fingerprint analysis, perhaps the most iconic forensic pattern recognition technique, as a case study, but most of the issues explored apply to all forensic pattern recognition disciplines.

Some key core competencies covered in this module include:

1. The probabilistic nature of most empirical knowledge.
2. The world as it actually is vs. the world as we discern it.
3. How lay intuitions about probability and statistics differ from the scientific understanding.
4. Decision making under uncertainty.
5. Concepts of validity.

AUDIENCE

We propose this module primarily for students of law and public policy, but it also may be of interest to journalism students concerned with science in the courtroom and criminal justice issues, and to students of criminology, science and technology studies, psychology and law, and other disciplines.

ORGANIZATION

The module is designed to be taught during 2 weeks of a standard professional-school course. It is divided into five "classes," which are expected to consist of around 90 minutes of in-class instruction time, plus additional time outside of class for reading and completing written assignments.

1. Validation
2. Admissibility
3. Reporting Results
4. Expertise and Policy Reform
5. Policy and Legal Reform

ACTIVITIES AND LEARNING OUTCOMES

While the module is adaptable to any pedagogical style, the module is formulated around an "active learning" approach. This approach is premised on the assumption that students learn better by doing activities than by passively receiving content. Each case is centered on a problem that is posed to the students. The precise classroom activities for problem solving could vary, according to instructor goals and preferences. Students could work individually or in a group; report back to the class at various intervals; engage in a variety of structured or unstructured discussions; write papers or give presentations.

OVERVIEW AND CONTEXT

Forensic Pattern Recognition in Historical, Scientific, and Legal Context

Definition

Forensic pattern evidence is a broad term for a wide variety of techniques that seek to make associations between physical traces. Usually the association of one or more traces can support an inference that might be helpful in the investigation of a crime. Fingerprinting is the most iconic form of forensic pattern evidence, but other types include firearms and toolmarks, bitemarks, handwriting, footwear and tire impressions, hair and fibers, and so on. Forensic pattern evidence can generally be distinguished from other types of forensic evidence in that it consists of the visual comparison and attempted association of images, impressions, or traces. Forensic disciplines that are *not* pattern evidence might include drug analysis, toxicology, arson and explosives, and medical examinations. Why these are *not* pattern evidence is not entirely clear. It does seem that these techniques are not comparative, in that they tend not to routinely compare one or more image, impressions, or traces. (There may be, however, a sense in which all techniques are comparative: For example, a claim that a substance is a drug does invoke an implicit comparison to some standard measurement for a known sample of that drug, even if the comparison is not done in every case.) Also important, is that some nonpattern evidence, such as drug analysis, uses instrumental analysis. Forensic pattern evidence overwhelmingly relies on analyses by human observers.

Forensic DNA profiling would seem to be somewhat of a special case. It undoubtedly takes the familiar form of a comparison of one or more traces. One the other hand, it relies upon instrumental analysis, albeit sometimes with some human interpretation added at the end of the process. It is probably for this reason that DNA profiling is often invoked as a model or foil for forensic pattern evidence.

History

Occasional uses of some forms of forensic pattern evidence, such as handwriting, date back centuries, but it would be fair to say that the routine use of forensic pattern evidence was a development of the 20th century. Due to a variety of historical circumstances, forensic pattern evidence acquired several general characteristics, including, as mentioned above, a reliance on the trained human observer as the comparative "instrument"; a reliance on nonquantified, subjective ways of characterizing observations, rarities, and so on; in some cases, such as fingerprints and firearms and toolmarks, very strong epistemological claims; a dearth of legal scrutiny of those claims; and a general location of the scientific work in law enforcement (the "crime laboratory" or the "identification unit").

It has now become commonplace to note how remarkably free of public criticism and legal scrutiny forensic pattern evidence remained for most of the 20th century. In the last two decades of the century, however, a growing number of observers began pointing out the idiosyncrasies of forensic pattern evidence. Chief among these, for our purposes, was the fact that the pattern recognition disciplines did not use numbers to quantify, for example, the rarity of the features being observed; the degree of similarity between known and unknown traces; or the accuracy rate of the technique. Instead, the pattern disciplines relied on semantic statements like "match," "consistent with," "in my expert opinion," "based on my training and experience," "reasonable degree of scientific certainty," "in my years of experience, I have never seen any two traces from different sources so similar," and so on. Forensic disciplines that did rely on numbers, such as serology and drug analysis and other chemical techniques, emerged early in the 20th century. However, it was the advent of forensic DNA profiling in the late 1980s that provided a stark, and often invidious, comparison with the pattern disciplines. During the 1990s, in a series of legal battles, colloquially dubbed "the DNA wars," lawyers and scientists engaged in heated disputes over how to calculate the rarity of a DNA profile and (to a much lesser extent) how to estimate the occurrence of laboratory error. As the DNA wars started to wane, it was shocking for many to discover that the pattern disciplines had "resolved" these questions simply by not attempting to estimate or quantify the rarity of features at all. Moreover, no one had even collected the data that would be necessary to make such estimates. And, courts had not required the disciplines to do either.

Further, at least two disciplines (fingerprints and firearms and toolmarks) found it acceptable to routinely make very strong testimonial claims of virtual or actual certainty. Such claims were also made in other disciplines, if not routinely. In the case of fingerprints, moreover, claims of "infallibility" and "zero error rate" were also common. In addition, the pattern recognition disciplines all lacked objective measurements, of the kind used in forensic DNA profiling, relying instead on subjective, "I know it when I see it" approaches. In some cases, these observations extended into outright criticism. This criticism, in turn, found its way into litigation, where defense attorneys were ethically obligated to challenge evidence that was offered against their clients. These challenges to forensic pattern evidence formed an important component of the now commonplace notion of a "crisis" in forensic science over the past two decades or so. (Other important components included challenges to forensic DNA profiling in the 1990s; criticism of other forensic disciplines such as medical examination, arson and explosives investigation, and drug analysis; crime laboratory scandals; and the innocence movement.)

The watershed moment in this crisis was undoubtedly the publication of the National Research Council report *Strengthening Forensic Science in the United States* in 2009. This report echoed many of the existing criticisms of forensic pattern evidence and lent them the imprimatur of the nation's leading scientific institution.

The unfolding of this crisis over the past two decades produced scientific, legal, and policy debates that generated the materials for this module. These debates often revolved around the question of whether various forms of forensic pattern evidence had been validated. Fierce legal battles were waged over the admissibility of various types of forensic pattern evidence. Numerous individuals and institutions have begun wrestling with the problem of how the results of forensic pattern analyses should be reported. And, as we write, a concerted effort to govern, regulate, and improve forensic science in the United States is being undertaken, principally by the United States Department of Justice and the National Institute of Standards and Technology, but also by organizations such as the National Academies, the American Association for the Advancement of Science, and the National Science Foundation.

Instructors' Guide for
Forensic Pattern Recognition Evidence

CLASS 1: VALIDATION

Learning Goals

- Students will be able to identify the scientific claim that is fundamental to the case argument.
- Students will develop strategies for evaluating empirical evidence offered in support of the appropriate scientific claim.
- Students will be able to distinguish between concepts of validity, reliability, precision, and accuracy, and will be able articulate how each influences the strength of a scientific claim.

Core Competencies Addressed

1. Scientific methods
2. Probability
3. Type I vs. type II errors
4. Validity
5. Reliability
6. The ideal of falsifiability and its limits
7. Peer review and publication/circulation of data and results
8. Testing hypotheses
9. Null hypothesis statistical testing
10. Randomized controlled trials

Activities

Students will first be asked to try to define the appropriate scientific claim that is being made on behalf of fingerprint identification. This will illustrate the notion of **scientific methods**. The goal is to prompt students to hone in on the **appropriate scientific claim** (e.g., fingerprint examiners can accurately determine the source of latent prints) and avoid seductive, but distracting, claims (e.g., all human friction ridge skin is unique; fingerprint identification is/is not "scientific"; fingerprint examiners "can" make accurate identifications; fingerprint identification follows the scientific method). Once they have established the appropriate claim, students will be asked to explore how this claim might be supported empirically, and practice evaluating such empirical information.

Suggested Class and Assignment Format

- In advance of class, students are assigned to read the Background Readings and complete a worksheet containing the following questions: What is the scientific claim made by fingerprint identification? What scientific information is necessary to answer this question? How might studies be conducted to provide this information? This will illustrate the notion of **scientific methods**.
- After completing the worksheet, but still in advance of class, students should complete the remainder of the readings.
- During class, the instructor should first go through the worksheet with the students for perhaps a third of the class period.
- The remaining portion of the class should be spent on Kafadar and Ulery. Ensure through discussion that students understand the key elements of Kafadar, and then apply them to Ulery.
- Students should chart out the false positive and false negative (type I and type II) errors to explore the relationship between them. This will illustrate the concept of **type I vs. type II errors.**
- Students should explore how the results of the Ulery study might be reported in a single short paragraph to a consumer of scientific information, such as a judge or policy maker.
- Students should critique the Ulery study and identify possible design flaws and areas for further research. This will illustrate the concepts of **scientific methods** and **peer review and publication/circulation of data and results.**

Readings

- Background readings:

    - Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, U.S. Department of Commerce, National Institute of Standards and Technology (2012).

      The purpose of this reading is to provide basic background information on fingerprint identification: where it comes from, what it purports to do, and how it purports to do it. From this reading, students should be able to identify the appropriate scientific claim that pertains to fingerprint identification. This will illustrate the concept of **scientific methods.** See especially pp. 1–9, which is an attempt to describe Analyze, Compare, Evaluate—Verify (ACE-V), the supposed methodology used in latent print analysis. (Whether *methodology* is an appropriate term has been questioned, but will be set aside here.) Students should identify that this is a "process map," and that describing the process does not answer the **appropriate scientific claim** that they have identified. This will illustrate the concepts of **scientific methods** and **validity.**

- E.J. Imwinkelried, “Coming to Grips with Scientific Research in Daubert’s ‘Brave New World’: The Courts’ Need to Appreciate the Evidentiary Differences between Validity and Proficiency Studies,” *Brooklyn Law Review* 61: 1247 (1995).

  The purpose of this reading is to introduce the classic scientific concepts of *validity*, *reliability*, *precision*, and *accuracy*. We want students to understand the technical distinctions between these concepts. This will illustrate the concepts of **validity, reliability,** and **the ideal of falsifiability and its limits.** We also want them to understand that courts often use them interchangeably. Finally, we want them to understand that in most cases, the relevant question for a court, policy maker, or other consumer of scientific information is that of accuracy.

- Examples of studies purporting to validate fingerprint analysis (summaries only, attached):

  - Francis Galton, *Finger Prints*, London: Macmillan, pp. 100–113 (1892).

    Galton attempted to estimate the probability of two exact duplicate sets of 10 friction ridge skin patterns existing in the human population. The probability he found was 1 in 4. This figure is often misreported as 1 in 64 billion, which was Galton’s estimate of the probability of any two *specified* sets of 10 friction ridge skin patterns being exact duplicates. The 1 in 4 figure is reached by dividing the 1 in 64 billion figure by the world population (in Galton’s time) of 16 billion fingers (on 1.6 billion people). It would be useful to help students recognize the distinction between two figures. (It is the “birthday problem,” well known in statistics.) However, the more important issue is to get students to recognize that the probability of duplication is not responsive to the **appropriate scientific claim**: how accurate fingerprint identification is. Even if duplication is zero, that only tells us about one potential source of inaccuracy (duplication). We still know nothing about another source of inaccuracy (error).
  - Kasey Wertheim, and Alice Maceo, “The Critical Stage of Friction Ridge Pattern Formation,” *Journal of Forensic Identification* 52 (1):35–85 (2002).

    Wertheim and Maceo begin their article by discussing the “validity” of latent print identification. However, the article does not discuss validity at all; it discusses the formation of friction ridge skin. If students correctly understand the definition, we hope that they will be able to understand that this article does not address it. This will illustrate the concept of **validity.** We also hope they will be able to understand why knowledge about the formation of friction ridge skin does not answer the **appropriate scientific claim** that they have identified. This fact should be clear without having read the entire paper, so it is appropriate for students to skim this paper.
  - Stephen Meagher, Bruce Budowle, and Donald Ziesig, “50K vs. 50K Fingerprint Comparison Test: Government Exhibits 6-8 and 6-9 in Pretrial Hearing,” *United States v. Mitchell*, Cr. No. 96-407, E.D. Pa (1999).

This study was introduced by the FBI in litigation in *United States v. Mitchell* in 1999. It notoriously concluded that the probability of nonduplication of a single finger friction ridge skin pattern was 1 in 10^{97}. Students should recognize the probability of nonduplication is not the **appropriate scientific claim** that they have identified.[1] This will illustrate the concepts of **scientific methods**, **validity,** and **probability.** Students may also note that the study was neither peer reviewed nor published, illustrating the concept of **peer review and publication/circulation of data and results.**

- Contemporary studies:

  - Karen Kafadar, "Statistical Issues in Assessing Forensic Evidence," *International Statistical Review* 83(1): 111 (2015).

    This article described an appropriate validation study that could be carried out for fingerprint identification. Students should note that it would measure accuracy. Students will be introduced to false positives and false negatives, and sensitivity and specificity, and the inverse relationship between them might be discussed. They will also be exposed to some considerations in study design, such as controls, random assignment, and so on. This will illustrate the concepts of **validity, type I vs. type II errors, testing hypotheses, null hypothesis statistical testing, randomized controlled trials,** and **the ideal of falsifiability and its limits.**
  - Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts, "Accuracy and Reliability of Forensic Latent Fingerprint Decisions," *PNAS* 108: 7733 (2011).

    This is one of the two earliest properly conducted accuracy studies of fingerprint identification. Students should recognize that, in contrast to the earlier studies, it does measure accuracy. Students should note that it reports results in a manner consistent with that suggested by Kafadar. Students should spend some time deciphering Figure 2, which is a particularly elegant graphical way of reporting the results of such a study. This will illustrate the concepts of **validity** and **type I vs. type II errors.**

    Some design flaws in the Ulery study include:

    1. The research subjects knew they were participating in the study, and their performance might not reflect their performance in actual casework. Students should be asked to debate how this flaw might

[1] It might also be useful if students identify design flaws in the study. In particular, the study compared Automated Fingerprint Identification System (AFIS) scores generated by comparing a friction ridge skin image to itself to AFIS scores generated by comparing a friction ridge skin image to images known to have come from other areas of friction ridge skin. This design does not appropriately model the situation in question. The proper design would be to compare AFIS scores generated by comparing a friction ridge skin image to a *different* image known to have come from the same area of friction ridge skin to AFIS scores generated by comparing a friction ridge skin image to images known to have come from other areas of friction ridge skin.

affect the results. Would the subject perform better in the study than in casework or worse? Which specific conclusions (individualization, exclusion, inconclusive) would they report more or less in the study compared with casework?

2. The population of subjects does not reflect the population of actual practitioners. Students should debate how specifically the subject population might be biased. The level of difficulty of the latent print materials and the proportion of comparisons that are same-source does not necessarily reflect actual casework. Identifying these design flaws will illustrate the concept of **peer review and publication/circulation of data and results**.

Assessment Questions [With Suggested Answers]

1. What's the difference between false positives and false negatives, and how can each impact a legal case differently in the context of a fingerprint comparison?

[**A:** A false positive occurs when a fingerprint examiner concludes that two fingerprints came from the same finger when they actually came from different fingers, while a false negative occurs when the examiner concludes they came from different fingers when they actually came from the same finger. In a criminal case, false positives can lead to investigation and even conviction of an innocent suspect, while false negatives can lead to elimination of a guilty suspect.]

2. If someone claims that fingerprinting is scientifically valid because there is scientific research explaining how fingerprints are formed, how can you respond? What if they claim the method is valid because it's been used in court for over a hundred years?

[**A:** Research explaining how fingerprints are formed does not show that an examiner is capable of recognizing when two fingerprints came from the same finger; the process of formation may lead to variation in friction ridge patterns, but does not demonstrate that such variation is likely to be viewable to an examiner, or that a human examiner is capable of accurately recognizing common features, or that the variation is sufficient to conclude that two prints came from the same finger. Use of a method in court for a long period of time says little about the state of the science, as our court system operates on a system of precedent, so courts are likely to admit a method as long as it has been previously admitted. Additionally, cross-examination in court does little to reveal an examiner's errors if the examiner does not know if she made an error, so claims of a history of accuracy in court are problematic.]

3. You're an attorney and opposing counsel is bringing in an expert at bite mark analysis to claim that your client bit into a piece of cheese found at a crime scene. You're preparing to cross-examine this expert and need to establish whether the

analysis of the mark in cheese is valid. What is the appropriate scientific claim for you to question? What kinds of studies would the expert need to present to demonstrate her analysis is valid?

[**A:** The expert is effectively claiming that she can accurately determine when an impression of teeth in cheese came from a particular suspect's mouth. Validation studies should demonstrate accuracy in this process, under circumstances similar to the ones in this case. An experimental design should, at a minimum, include a reasonably large number of comparisons of tooth impressions by bitemark experts, and differentiate between false positive and false negative errors. Students do not need to create an in-depth study design, but should recognize the importance of an objective assessment of accuracy in comparisons.]

CLASS 2: ADMISSIBILITY

Learning Goals

- Students will gain an understanding of the various legal approaches to evaluating expert testimony and the admissibility of forensic pattern evidence.
- Students will be able to evaluate the Frye, Daubert, and the no-threshold approaches, and to assess the justifications, rationales, strengths, and weaknesses of each.
- Students will be able to articulate the social basis for scientific reputations, and to assess the extent that reputations can be assessed and measured by lay outsiders.
- Students will be able to evaluate the effects of litigation on scientific knowledge production.

Core Competencies Addressed

1. Paradigm-shifting visionaries vs. cranks
2. Resolving scientific disputes
3. Lay v. expert
4. Scientists' involvement in litigation
5. Issues of responsibility, expertise, and credibility
6. The ideal of falsifiability and its limits
7. Publication bias
8. Science in the courts
9. Admissibility
10. Daubert, etc.
11. Prosecutor's fallacy
12. Nature of scientific consensus
13. The contingent nature of scientific standards and findings

Activities

Students will be asked to consider and weight the relative merits of the Frye, Daubert, and no-threshold approaches. In groups, they will then apply each approach to a hypothetical proffer of expert testimony. Finally, students will be asked to consider how the admissibility determination should be conducted, and to recommend an admissibility standard for a jurisdiction that is writing a new evidence code. This will illustrate the core competencies **science in the courts, admissibility, resolving scientific disputes, scientists' involvement in litigation,** and **Daubert etc.**

Suggested Class and Assignment Format

- In advance of class, complete all readings.
- Discuss the relative merits of the *Frye*, *Daubert/Kumho Tire,* and no-threshold approaches.
- Assign students in groups to the three approaches.
- Give each group the provided proffer of expert testimony on fingerprint identification, and give them time to make and support an admissibility decision.
- Ask each group to give detailed answers to the components of their admissibility standards (e.g., how is the "relevant scientific community" constituted? what is meant by "error rate"?).
- Students can then be asked, individually, in groups, or as a whole, to recommend an admissibility standard for a jurisdiction that is writing a new evidence code.

Readings

- Legal standards and approaches:

    - Federal Rule of Evidence 702.
    - *Frye v. United States*, 293 F 1013 (D.C. Cir. 1923).

      The case concerning the admissibility of lie detector test established the earliest "admissibility standard" in U.S. law. This will illustrate the core competencies **science in the courts** and **admissibility.** The standard it established was "general acceptance in the community to which it belongs." Students should understand that this is a "deference" model—the judge assumes her own incompetence to judge the merits of the science, instead deferring to the scientific community by attempting to assess its "general acceptance" of the methods. This will illustrate the concepts of **lay v. expert** knowledge and **issues of responsibility, expertise, and credibility**.

      Students should explore the advantages of this approach—untrained judges do not play "amateur scientist." They should also explore the disadvantages: First, that cutting edge, but valid, science might be excluded for lacking general acceptance. This illustrates the concept of **publication bias,** the **nature of scientific consensus,** and **the contingent nature of scientific standards and findings.** Second, that scientific knowledge claims produced by closed communities, cults, or what Adina Schwartz calls "mutual admiration societies" might be admitted, despite lacking validity. This will illustrate the concept of **paradigm-shifting visionaries vs. cranks.**

In addition, it is unclear how "general acceptance" is defined, or what constitutes the "community in which it belongs," which has come to be reformulated as "the relevant scientific community." Students should understand that the judge's choice of "community" can determine the outcome of the admissibility inquiry: if the community relevant to the admissibility of a lie detector is a group of lie detector operators, it is likely to be ruled admissible because such a community is obviously accepting of the science of lie detection. If the community consists of experimental psychologists, who are less accepting of its validity, the science is less likely to be ruled admissible.

- *Daubert v. Merrell Dow Pharmaceuticals*, 507 U.S. 579 (1993).

  This landmark scientific evidence case found that the Federal Rules of Evidence superseded *Frye* in federal courts, and was subsequently adopted, or imitated, in more than half of U.S. jurisdictions and some foreign countries. It imposed on federal judges a gatekeeping responsibility for scientific evidence. Rather than asking only whether the relevant scientific community has accepted a science, the judge now takes on questions about whether the field or method is falsifiable, whether it has been sufficiently tested, its error rate, and whether the method has been subject to peer review. This will illustrate the core competency **Daubert etc.** *Daubert's* reference to falsifiability will illustrate **the ideal of falsifiability and its limits.**

  Students should recognize the major change from *Frye*: the judge herself is now responsible for assessing a method's validity, based on a number of criteria relevant to scientific testing. During their class activity, students should question whether the state of research in fingerprint methods (as studied in Class 1) would qualify for admissibility based on the *Daubert* standard. *Herrera* (below) provides another opportunity for this discussion.

- *Kumho Tire Company, Ltd. v. Carmichael*, 526 U.S. 137 (1999).

  This opinion is an important counterpart to *Daubert*, extending the *Daubert* inquiry to all expert testimony (as opposed to only that labeled "scientific"), and recognizing the breadth of expertise likely to be recognized in court. It is important to note that it declines to distinguish between "scientific" and "technical" expertise. This case also is significant because it further defines what counts legally as a "reliable" method, and underscores the basic scientific assumptions that underlie courts' assessment of even nonscientific methods. This illustrates the concept of **lay v. expert** knowledge.

- *United States v. Herrera,* 704 F.3d 480, 482 (7th Cir. 2013).

This is a 2013 Seventh Circuit decision evaluating the admissibility of fingerprint evidence under the *Daubert/Kumho* standard. We hope that students' own evaluations during the class exercises will differ from this opinion, which assumes that professional certification, the "appearance" of a lack of error, and training of examiners equate to a measurement of reliability.

There are also a number of errors in statistical reasoning in *Herrera* that the instructor may wish to draw out. It misinterprets Galton's estimation of the probability of duplication (discussed in Class 1). It commits the **prosecutor's fallacy**, equating the probability of evidence being spurious with the probability that the defendant "should be acquitted." And, it commits the "source probability error," which involves equating the random match probability with the probability that someone other than the "matchee" is the source of the matching DNA profile.

- *Commonwealth v. Patterson,* 840 N.E.2d 12 (Mass. 2005).

  This is a 2005 decision by the Supreme Judicial Court of Massachusetts evaluating a specific form of fingerprint evidence ("simultaneous impressions") under *Lanigan,* a *Frye*-like standard. We hope that students' own evaluations will differ from this one. In particular, we hope they will not limit the "relevant scientific community" to professional latent print examiners, and that they will not so easily dismiss the evidence of an entrenched orthodoxy within that community. This will illustrate the concepts of **lay v. expert** knowledge and **issues of responsibility, expertise, and credibility**.

- (Optional) Reading to facilitate students' analysis of legal opinions:

  - Orin S. Kerr, "How to Read a Judicial Opinion: A Guide for New Law Students."

    Although addressed to law students, this essay provides a helpful guide to anyone unfamiliar with judicial legal opinions. It explains what judicial opinions are, how they are structured, and what students should look for when reading them.

Assessment Questions [With Suggested Answers]

1. Under the *Frye* standard, what are some potential pitfalls for courts evaluating forensic science? Under *Daubert*?

[**A:** The *Frye* standard creates difficulties in defining the "relevant scientific community" when it comes to forensic science, because the most obvious community is the community of forensic practitioners, whose livelihood depends on their methods being admissible in court and who are selected for their acceptance of the method (i.e., they are in that community—they became forensic scientists—*because* they accepted its validity, so the court would automatically ignore any scientists who *don't* accept the method). The *Daubert* standard has tripped up courts who mistake one examiner checking the conclusions of another for peer review, and years of use of forensics in court as testing. There are other possible answers to this.]

2. If you were a lawyer arguing against admissibility of fingerprint evidence, what evidence might you introduce during a *Daubert* hearing? If you were arguing for its admissibility, what evidence might you introduce?

[**A:** Evidence against admissibility might include an expert (or papers by experts) outside the field, such as academics, explaining the limitations of existing studies of fingerprint validity. Evidence for admissibility might include more recent studies, such as Ulery's from Class 1.]

3. What would you consider the "relevant scientific community" if you were a judge evaluating a different type of pattern evidence under *Frye*, such as shoe impressions?

[**A:** This question is a bit tricky; students should be able to recognize the difficulty of defining the proper community, and realize that the community of shoe-impression experts, who by definition have accepted the validity of shoe-impression methods, is not helpful in determining the validity of a method.]

CLASS 3: REPORTING I (HYPOTHESIS TESTING)

Learning Goals

- Students will be able to understand and apply key probability principles, including uncertainty, frequencies, the product rule, and conditional probabilities.
- Students will be able to recognize that all scientific claims involve uncertainty and to develop strategies for quantifying that uncertainty.

Core Competencies Addressed

1. Certainty vs. uncertainty
2. Methods
3. Correlation vs. causation
4. Measurement errors
5. Reliability
6. Testing hypotheses
7. Null hypothesis statistical testing
8. Randomized controlled trials
9. Likelihood ratio approaches
10. Bayes' Theorem

Activities

Students will be asked to evaluate and critique a proposed conclusion and reporting standard and explain its shortcomings, given the current state of scientific knowledge.

Suggested Class and Assignment Format

- Readings should be completed in advance of class.
- Instructors should first emphasize the key points to be taken from the Aitken et al. reading (see below).
- The provided exercises will ensure their understanding of key principles, such as the importance of understanding and quantifying uncertainty, how to calculate frequencies, the product rule, and conditional probabilities. This will illustrate the concepts of **certainty vs. uncertainty, methods, correlation vs. causation, measurement errors, reliability, testing hypotheses, null hypothesis statistical testing,** and **randomized controlled trials.** Instructors should use the provided problems however they see fit: divide the class into groups to solve them, assign them as a worksheet, work through them together in class, and so on. Instructors should make sure students understand the reasoning involved in each problem before moving on.
- Students will then evaluate whether the 2003 SWGFAST reporting standard comports with the principles articulated in the reading and with what they learned from the provided exercises. This may be expressed as having two components:

    - How helpful are these standards for examiners in determining what conclusion to reach during a fingerprint comparison?
    - Is the resulting conclusion the best way to present fingerprint evidence in court? Why or why not?

Readings

- Key probability concepts:

    - Colin Aitken, Paul Roberts, and Graham Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*. (Focus in particular on 1.3–1.5, 1.14–1.22, and Chapter 2; skim Chapter 3).

      Probability is a way to use math to conceptualize uncertainty and apply it to reasoning or decision making. This will illustrate that there is no such thing as scientific certainty—that all scientific claims involve uncertainty and that it is desirable to quantify uncertainty. Probabilistic presentation of results can allow this, while categorical, all-or-nothing statements are unhelpful and deceptive. This section will illustrate the concept of **certainty vs. uncertainty**.

The most important thing to take away from this class about uncertainty is that it's inevitable. Nothing in science is absolute, and no result is ever 100% certain. This does not mean, however, that we know nothing, or that decisions cannot be made based on information simply because it isn't certain. Instead, it's important to decide how much uncertainty is acceptable, and to determine how much uncertainty there is in a statement. In the case of forensic fingerprinting, this means that no matter how good the fingerprint examiner, or how clear the fingerprints, no conclusion of "match" or "nonmatch" is ever absolutely certain; and most fingerprint comparisons are less than ideal, so some conclusions have more uncertainty than others. To draw any conclusions from a fingerprint comparison, such as a jury's verdict, the amount of uncertainty is important.

Instructors should ensure that students understand the methods necessary to work through the provided problems; focus in particular on those parts of the reading pertaining to frequency, likelihood, **likelihood ratios,** and **Bayes' Theorem**. A brief explanation of some important probability concepts can be found in the Glossary accompanying this module, and a closer read of Chapter 3 might be useful for instructors anticipating student confusion.

- Reporting standards:

  - SWGFAST Scientific Working Group on Friction Ridge Analysis Study and Technology, Standards for Conclusions (Draft for Comment) (2003)

    This is a guideline issued by a standard-setting body for latent prints in the United States; that is, it provides latent print examiners with standards for what conclusions they can draw from a comparison, and how to decide between them. The standards are minimal, providing only three options (one of which is "inconclusive"). Students should recognize that the guideline is categorical, rather than probabilistic, and, therefore, inconsistent with the principles of statistical reasoning articulated in the previous reading. In other words, it makes a claim of certainty, rather than seeking to quantify the uncertainty associated with the claim. They should recognize the lack of guidance for examiners deciding when to reach a particular conclusion, and that it is problematic to simply tell a court the comparison results in an identification or an exclusion, without more—that *some* probabilistic expression of the strength of the conclusion is necessary.

Assessment Questions [With Suggested Answers]

1. Why is it important for a forensic examiner to produce a result in a probabilistic format rather than a categorical conclusion?

[**A:** Uncertainty always exists, but can vary widely. One fingerprint "match" might be more certain than another, by containing more features and having more clarity in the impression, and this variation should be available to fact finders who must make a decision based on the evidence. A categorical formulation of a comparison as matching or non-matching deceives the fact finder into believing that the conclusion is certain.]

2. Assuming an expert witness has found a way to quantify the uncertainty about her conclusion—that is, she has come to a probabilistic conclusion—should that quantification be conveyed to the fact-finder (judge or jury)? Why or why not? If yes, how?

[**A:** So long as the student makes a compelling argument, that answer will be correct. Some understanding of the degree of uncertainty should probably be conveyed to the fact finder, but whether this should be described quantitatively (using numbers) or in qualitative phrases is an open question. It depends on what would be useful to a judge or jury.]

3. Most standards for conclusions for fingerprint examiners, like the SWGFAST version, give examiners the option of reaching a result of "inconclusive." Does this solve the problem of categorical conclusions? Why or why not?

[**A:** Inconclusive is not a solution to the problem of categorical conclusions because it is still categorical. Even if an examiner has some idea of a cutoff for certainty, whereby she only concludes something is a "match" if she's reached a particular degree of certainty, that "match" is still not absolutely certain and that uncertainty should still be understood in a probabilistic way.]

CLASS 4: REPORTING II (DECISION THEORY AND LIKELIHOOD RATIOS)

Learning Goals

- Students will understand the key features of decision theory and likelihood ratios.
- Students will be able to recognize common fallacies in interpreting probabilities, including the prosecutor's fallacy.
- Students will be able to evaluate proposed reporting standards and to assess the strengths and weaknesses of these standards as guidance to fingerprint examiners who testify in court.

Core Competencies Addressed

1. Likelihood ratio approaches
2. Prosecutor's fallacy
3. Decision theory

4. Bayes' Theorem
5. Combining information
6. The idea of priors
7. A Bayesian approach to testing hypotheses
8. Biases in perceiving, remembering, and analyzing information (JDM and social psych)
9. Lay intuitions about probability and statistics
10. Decision making under uncertainty

Activities

Students will be asked to evaluate and critique various proposed reporting standards and make a recommendation for how the results of latent print analyses should be reported in court, given the current state of scientific knowledge.

Suggested Class and Assignment Format

- All readings should be completed in advance of class.
- Instructors will then ensure that students understand the key features of **decision theory**.
- Once the nature of "utility function" is understood, the instructor should engage a discussion of who should assign the utility function in a legal setting—the expert or the fact finder—a question upon which Biedermann et al. (at least in this article) purport to be agnostic.
- Students will then evaluate whether the 2011 SWGFAST reporting standard comports with the principles articulated by Aitken et al. and with **decision theory**.
- Students will then discuss the Neumann article. The instructor will endeavor to ensure they understand the concepts of rarity and **likelihood ratios** (LR). Students should understand the **prosecutor's fallacy** (see Glossary). Students should also discuss whether, if it is possible to make rarity estimates for all pattern evidence, it is acceptable *not* to do so.
- Students will then evaluate whether the U.S. Army Crime Laboratory reporting standard comports with the principles articulated by Aitken et al. and with **decision theory**. They will discuss the weaknesses of the standard.
- Students should draw on their understanding of **uncertainty** and common **fallacies** to discuss how the various reporting standards will be understood by fact finders, such as juries or policy makers. This will illustrate the concepts of **biases in perceiving, remembering, and analyzing information (judgment and decision making and social psychology)** and **lay intuitions about probability and statistics**.
- By the end of class, either individually, in groups, or as a class, students should put together a recommendation for new reporting standards. Students should pay particular attention to the utility function and the issue of who should assign it.

Readings

- Key probability concepts, including the **likelihood ratio**, **prosecutor's fallacy**, and **decision theory**:

  - Alex Biedermann, Silvia Bozza, and Franco Taroni, "Decision Theoretic Properties of Forensic Identification: Underlying Logic and Argumentative Implications," *Forensic Science International* 177: 120–132 (2008).

    This article explains **decision theory** and explores how it could be applied to identification evidence, such as fingerprint evidence. Key points to note are the problems with categorical reports; the fact that each of the two hypotheses must have non-zero probability; that estimates of the probability of each hypothesis should be integrated with the "utility function" (the decision maker's preferences), and the fact that conclusions of "individualization" are not necessary for identification to be useful. This will illustrate the concept of **decision making under uncertainty, Bayes' Theorem, combining information,** and **the idea of priors.** Put another way, *certainty* is not necessary for identification to be useful.

    Decision theory is a framework for understanding the best decision to make given certain circumstances—in this case, given the particulars of a fingerprint comparison, the degree of uncertainty, the potential consequences of error, and so on. The utility function allows the examiner to incorporate the potential consequences of errors into that decision calculation, even recognizing that false positives have different consequences (potential for false conviction) than false negatives (potential for falsely eliminating a suspect).

    This is a highly technical article. Students do not need to try to follow the equations. However, they should try to grasp what decision theory is and how it applies to identification evidence. It is important that students try to grasp the "utility function." In addition, students should try to grasp what the authors are arguing and how it differs from other ways of conceptualizing forensic identification evidence that they have encountered. Students can probably skip Section 4 of this article.
  - Cedric Neumann, "Fingerprints at the crime-scene: Statistically certain, or probable?" *Significance* (Feb. 2012): 21.

    This article from the American Statistical Association's magazine intended for a lay audience is meant to be a popularized version of Neumann et al.'s seminal paper in the *Proceedings of the Royal Statistical Society*; it lays out a probabilistic approach to latent print evidence using **likelihood ratios**. This will illustrate **a Bayesian approach to testing hypotheses.** A key point is the need for rarity estimates of features found in common between two fingerprint impressions. On this point, Neumann et al., Aitken et al., and others of this school disagree with Kafadar (see Class 1), who adopts a more "black-box **validation**" approach. A second key point is that it is at least theoretically possible to make rarity estimates for pattern evidence like fingerprints, even if it is more difficult than it was for DNA profiles. This possibility should raise the question of whether it is any longer acceptable, from a legal or policy standpoint, not to do so. Again, this article dismantles the shibboleth of *certainty*, replacing it with a probabilistic understanding of pattern evidence.

- Existing and proposed reporting standards:

  - SWGFAST, Scientific Working Group on Friction Ridge Analysis Study and Technology, Standards for Examining Friction Ridge Impressions and Resulting Conclusions (2011).

    This is a revision of the 2003 guidelines from the previous reading. We hope that students will perceive that it recognizes the existence of probability in a way that the prior version did not. Nonetheless, we hope that students will perceive that it still does not fully adhere to the principles of statistical reasoning or of decision theory. In particular, it does not really situate the probability of the alternate hypothesis between 0 and 1—it is essentially 0—and it assumes the value of this probability, rather than counseling the analyst to estimate it. It offers no guidance as to how the probabilities are to be estimated or assigned, and it violates decision theory by ignoring the utility function. Again, it is a claim of *certainty,* rather than uncertainty.
  - Department of the Army Defense Forensic Science Center Information Paper: Use of the term "Identification" in Latent Print Technical Reports (November 3, 2015).

    This is the first statement by a U.S. crime laboratory acknowledging the problems with the terms "identification" and "individualization" and showing that these terms are not necessary for fingerprint evidence to be useful in criminal investigations. Students should recognize that the statement seeks to follow the principles of **probabilistic reasoning**. However, we hope students will also recognize some of the weaknesses of the statement. Most importantly, the phrase "extremely low" is vague, and the calculations or reasoning process that produced it are not explained.

Assessment Questions [With Suggested Answers]

1. An expert testifies that a DNA profile found at the scene of a crime is 1 million times more likely if the defendant was the source of the DNA than if someone else was the source. If a juror commits the prosecutor's fallacy, what does she *think* the expert is saying?

[**A:** A juror committing the prosecutor's fallacy would think the expert is saying that the defendant is 1 million times more likely to be the source of the DNA than not, given this DNA profile.]

2. According to Bayes' Theorem, what information does one need, in addition to the

likelihood ratio, to calculate the odds of a hypothesis being true given the evidence?

[**A:** The prior odds—the odds of the hypothesis being true *before* the evidence is incorporated.]

3. What is the difference between the black-box validation approach to fingerprint analysis taken by Kafadar in Class 1 and the Bayesian approach proposed in Class 4?

[**A:** The black-box validation approach looks for a general measure of accuracy in comparisons regardless of the frequency of particular features of the fingerprints involved. The latter approach looks to how common a fingerprint pattern might be in the population to determine its evidentiary strength.]

CLASS 5: POLICY & LEGAL REFORM

Learning Goals

- Students will develop a critical approach to assessing the relationship between science, policy, and law.
- Students will be able to articulate the American legal system's approach to expert testimony and to evaluate the merits and deficiencies of alternative approaches.
- Students will be able to articulate some of the differences between forensic science communities and other scientific communities, and to assess why such differences might be meaningful in formulating legal and/or policy interventions.
- Although not a specific subject of discussion, we hope that by discussing this issue in this way, students will recognize that they will find a **nuanced understanding of the nature of scientific knowledge** most useful in generating legal and/or policy interventions.

Core Competencies Addressed

1. Scientists' involvement in policy
2. Who sets standards
3. Role of learned societies
4. Role of NSF/NIH/other research agencies
5. Science Advisor/OSTP
6. Self-regulation
7. External regulations
8. What constitutes a scientific consensus and implications for policy decisions
9. State of forensic science

Activities

Students will be asked to generate a set of policy recommendations to: improve forensic science; better integrate "forensic science" with what is conventionally called "mainstream science"; improve the handling of forensic science by the courts; facilitate the use of societal resources (e.g., government, educational, institutional) to improve forensic science; generate a plan for the long-term regulation and governance of forensic science. These activities will illustrate the concepts of **scientists' involvement in policy,** and **what constitutes a scientific consensus and implications for policy decisions.**

Suggested Class and Assignment Format

- All readings should be done in advance of class
- Based primarily on the Thompson reading, students will first be asked to debate whether the problems with pattern evidence are best addressed by legal interventions, policy interventions, or both.
- Next, students will be asked to design appropriate policy and/or legal interventions. Students will be asked to critique one another's proposals and debate the pros and cons of all proposed interventions. While some proposals may be found in various readings, students should also be encouraged to develop their own.
- Finally, students will be asked to consider and propose legal reforms. What are the advantages and disadvantages of the commonly proposed reforms? Can the students think of other reforms? The instructor might want to refer back to Class 2 and recall the rationales for the various (*Frye, Daubert*, admit everything) standards.

Readings

- Who should reform pattern evidence? The government? The courts? The market?

    - William C. Thompson, "The NRC's Plan to Strengthen Forensic Science: Does the Path Ahead Run through the Courts?" *Jurimetrics* 50: 35 (2009).

      This article provides a useful summary of the debate over whether reform of forensic science in general (including pattern evidence) is best pursued through legal or policy interventions. Students should learn that the NRC Report, citing the "utter failure" of the courts to adequately regulate forensic science, favored government intervention, proposing the creation of a new government agency for regulating forensic science. Thompson, while not disagreeing about the "utter failure," contends that the courts are too influential to be ignored and forensic science will not be reformed unless courts take their regulatory responsibilities more seriously. This will illustrate the concepts of **who sets standards,** the **role of learned societies,** the **role of NSF/NIH/other research agencies, Science Advisor/OSTP,** and **external regulations.**
    - Roger Koppl, "How to Improve Forensic Science," *European Journal of Law and Economics* 20: 255 (2005).

This article explores a market approach to improving forensic science and makes a number of specific proposals to resolve the problems and bias issues caused by "epistemic monopolies." A crime lab typically has a monopoly on analysis of evidence, in contrast to other types of scientific analysis that involve competition between researchers, and is tied closely to police and prosecutors. Information sharing with police can lead to unconscious bias on the part of examiners, while the lab's dependence on prosecutors and police for work can lead to a pro-prosecution bias, even among well-intentioned examiners. Additionally, an absence of quality control and competition can allow error and fraud to go unchecked. Students should discuss the potential for the article's proposals, including "rivalrous redundancy" (wherein some evidence is sent to multiple labs) and independence from prosecutors and police. However, they should also note a number of practical problems raised by Koppl's proposal, including the costs of redundancy.

- Understanding forensic science:

    - Simon A. Cole, "Acculturating Forensic Science: What Is 'Scientific Culture,' and How Can Forensic Science Adopt It?" *Fordham Urban Law Journal* 38: 435 (2010).

      This article argues that the models of "science" cherished by many philosophers and scientific institutions are unlikely to be very helpful in thinking about how to make forensic science more "scientific." It suggests, instead, thinking about the nature of forensic science as an activity and the various tasks it is called upon to perform. It suggests dividing the forensic enterprise into a variety of subtasks, such as research, evidence collection, lab management, analysis, and interpretation, and designing an appropriate "scientific culture" for each one.

      Under this argument, only the research subtask is best served by the traditional model of science as a process of skeptical hypothesis testing; other subtasks should have other priorities, and therefore different types of reform. Students should understand the general argument that different tasks may require different "scientific cultures," and be able to understand or formulate a normative argument for particular cultural priorities for a given task.
    - Jennifer L. Mnookin et al., "The Need for a Research Culture in the Forensic Sciences," *UCLA L. Rev.* 58: 725 (2011).

      This article nicely summarizes the state of forensic science, with a particular emphasis on pattern evidence, and makes some quite reasonable proposals for improvement. It illustrates the **state of forensic science, who sets standards,** the **role of learned societies, self-regulation,** and **external regulation.** Students should be able to incorporate some of the observations about the current state of forensic science, and potentially some of the proposals for improvement, into their discussion of potential reforms.
    - David Sonenshein and Charles Fitzpatrick, "The Problem of Partisan Expert and the Potential for Reform through Concurrent Evidence," *Rev. Litig.* 32: 1 (2013).

This article restates the problem of expert evidence (in a civil law context), and describes some proposed solutions. The article reviews common "legal" reform proposals—proposals to improve pattern evidence via legal process, rather than via interventions by government, markets, or scientific institutions. Proposals include specialized "science courts" consisting of expert fact finders, court-appointed experts (who would, according to the proponents of this reform, lack the bias of experts hired by one side of litigation or the other), and concurrent evidence. Concurrent evidence, a method sometimes called "hot-tubbing" used in Australian courts, consists of experts from both sides testifying simultaneously and having a discussion about the evidence.

The purpose of this article is not necessarily to advocate for these reforms (the article contains its own critique of science courts and court-appointed experts; for a skeptical view of concurrent evidence, see Gary Edmond, "Merton and the Hot Tub: Scientific Conventions and Expert Evidence in Australian Civil Procedure," *Law & Contemp. Prob.* 72: 159 (2009), http://www.jstor.org/stable/40647170). Rather, it is to make students aware of common reform proposals so that they do not "reinvent the wheel." It is hoped that students will critically evaluate common reform proposals and possibly come up with innovative proposals of their own.

Assessment Questions [With Suggested Answers]

1. What are some arguments for and against reforming forensic science via the court system and legal evidence standards, as opposed to using direct government interventions into the policies of forensic institutions?

[**A:** Arguments for legal standards might include effectiveness because forensic scientists rely on their work being used as evidence and so would be motivated to comply with reforms, or protection of defendants because violating a court's standard would have direct consequences to the case. Arguments against legal reform include the fact that it has yet to occur—years after *Daubert*, courts still routinely accept a wide variety of forensic sciences that arguably lack sufficient scientific foundation. Courts also may not be able to create more nuanced reform.]

2. What are some differences between forensic science communities and other scientific communities? Why might these differences cause problems?

[**A:** The article by Mnookin and colleagues contains a number of examples of such differences and the problems they can cause, so a student could name a number of them. The overarching theme is the idea of a research culture; forensic scientists are not trained or treated as scientists, with a skeptical approach and a policy of routinely questioning methods, conclusions, and so on. This creates a lack of evolution in forensic science even when the foundational knowledge evolves, and means that many forensic scientists may not be aware of current thinking and may continue with outdated methods. They may also be susceptible to outside influences and biases that other types of scientists are more equipped to guard against by using blind study methods. Forensic scientists also tend to work directly for, or closely with, police and prosecutors, creating a culture of dependence on the prosecution side of every case and preventing free flow of information.]

3. Describe some of the potential problems with partisan expert evidence. Given these problems, why might this system have stuck around in American courts? When proposals such as specialized science courts and court-appointed experts have been on the table for years, why do we still have the current system for most cases?

[**A:** Some of the problems with proposed solutions are cost-based. If a court-appointed expert is used, who pays for it? Others are practical difficulties: how does a court choose an appointed expert? What prevents the court from appointing an expert who holds unusual views in the field? How is it decided what cases are heard by specialized courts—if only some of the evidence is scientific, or if the evidence includes multiple specializations, what court should hear it? Courts may also have not adopted these proposals widely simply due to tradition (the U.S. an adversarial system, after all) or difficulty with the transition. Students can answer this question with a number of theories, so long as they engage meaningfully with the question.]

Glossary

Fingerprint Terms

- ***Friction Ridges:*** The ridges that make up the patterns used in fingerprint examination.
- ***Friction Ridge Skin:*** The skin on which friction ridges are found; the palms of hands and soles of feet.
- ***Fingerprint:*** The pattern formed by friction ridges on the fingertip, which may be transferred onto other surfaces via oils, inks, 3-dimensional impressions, or other transfer methods.
- ***Latent Fingerprint:*** A fingerprint left on a surface accidentally, usually in sweat and oil from the skin, that cannot be readily discerned by the naked eye without enhancement, usually by fingerprint powder; the term is often used to refer to any fingerprints left at a crime scene, to differentiate from known fingerprints.
- ***Known Fingerprint:*** A fingerprint taken directly from its source, so it is known what finger it came from; includes any fingerprints taken by police or other entities, and stored with identifying information about the person and finger they came from (when comparing fingerprints, the question is usually whether there is a match between known and unknown samples—that is, the print from the suspect and the print from the crime scene).
- ***ACE-V ("Analysis, Comparison, Evaluation, Verification"):*** The process taught to fingerprint examiners for comparing two fingerprints.
- ***SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study and Technology):*** An organization created to establish standards for fingerprint examiners until the formation of the Organization of Scientific Area Committees (OSAC).
- ***Loop, arch, and whorl:*** The three general categories of fingerprints, based on the overall flow of friction ridges on the finger.
- ***Minutiae:*** Distinctive details in friction ridge patterns that can be used to compare prints; two common minutiae are ridge endings (where a friction ridge comes to an end) and bifurcations (where it splits into two ridges).
- ***Identification/Individualization:*** One of three conclusions used by fingerprint examiners; expresses the examiner's opinion that two fingerprints share sufficient traits that they came from the same finger.
- ***Exclusion:*** One of three possible conclusions used by fingerprint examiners; expresses the examiner's opinion that two fingerprints did not come from the same finger.
- ***Inconclusive:*** One of three possible conclusions used by fingerprint examiners; expresses the examiner's opinion that there is insufficient information to reach either an identification or exclusion conclusion.

Probability Concepts

- ***Correlation vs. Causation:*** When one variable changes whenever a second variable changes, the two variables are correlated. This can occur because one variable's change *causes* the other to change; this would be causation. But a correlation can result from other things, such as a third variable affecting both variables together, or causation running in the opposite direction from the one first assumed. Many people confuse correlation and causation, concluding causation from correlation evidence without realizing there may be other explanations for a correlation.
- ***Frequency:*** In the context of pattern evidence, frequency usually refers to how many of a pattern or feature of the evidence exist in the general population; that is., how common it is. If a fingerprint shows the general pattern of a whorl, we might ask how many fingers have whorl prints in the population; this is important because if whorls are common, then having a whorl on a crime scene print and on a suspect's finger means less than if whorls are rare.
- ***Conditional Probability:*** A probability of one thing occurring if a second thing is true. For instance, the probability of a random fruit in a grocery store being a banana might be fairly low because grocery stores carry many types of fruit, but the *conditional* probability of a fruit being a banana *if we already know that the fruit is yellow* (the condition) is likely to be much higher. This probability takes the following form: P(banana|yellow), the probability of something being a banana if we know it's yellow. Note that this is a different probability from the chance of something being yellow if we know it's a banana, or P(yellow|banana). A banana may be one of many yellow things, so that P(banana|yellow) is small, but most bananas are yellow, so P(yellow|banana) is high. It's important to avoid confusing these two probabilities, which is a very common mistake.
- ***Likelihood Ratio (LR):*** A ratio of two conditional probabilities. It's generally used to evaluate how strongly a piece of evidence supports one theory as opposed to another. LRs are commonly used for DNA evidence, which looks like this: P(DNA profile|suspect left DNA)/P(DNA profile|someone else left DNA). This can be stated as the ratio between the likelihood of this DNA profile in the crime-scene DNA if the suspect left it vs. if someone else did. The higher the ratio, the stronger the evidence—if this evidence is 5 times more likely to appear if the suspect is guilty than if they're innocent, it's evidence of guilt, but if the evidence is 1 trillion times more likely to appear if they're guilty than if they're innocent, that is much stronger evidence of guilt.
- ***The Prosecutor's Fallacy:*** When the mistake described under "conditional probability" above is made about the likelihood ratio from evidence. Someone making this mistake sees "this evidence is 100 times more likely if they're guilty than if they're innocent," and thinks it means "they're 100 times more likely to be guilty than innocent." The first statement is the odds of the evidence given guilt, while the second is the odds of guilt given the evidence. The difference can be difficult to spot at first, which is why this mistake is so common.

- ***Bayes' Theorem:*** A mathematical way to calculate the odds of guilt (or some other theory) given evidence, using the LR for the evidence. Note that this is what someone making the mistake of the prosecutor's fallacy believes the LR gives us—Bayes' Theorem uses the LR to calculate this, but requires more information. Specifically, you need the odds of guilt without the evidence; Bayes' is a way of changing those odds when new evidence is incorporated into the calculation.

READINGS AND EXERCISES
(CLASSES 1–5)

CLASS 1

Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice Through a Systems Approach*. U.S. Department of Commerce, National Institute of Standards and Technology (2012), 1-9. Document in public domain.

Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach

The Report of the Expert Working Group on Human Factors in Latent Print Analysis

In Memoriam
This report is dedicated to the memory of Danny Greathouse, a valued contributor to this study and a friend who will be missed.

Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach was produced with funding from the U.S. Department of Justice's National Institute of Justice and in collaboration with the Law Enforcement Standards Office in the U.S. Department of Commerce's National Institute of Standards and Technology.

Suggested citation format: Expert Working Group on Human Factors in Latent Print Analysis. *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. U.S. Department of Commerce, National Institute of Standards and Technology. 2012.

Table of Contents

Chapter 1: The Latent Print Examination Process and Terminology

Introduction

The conventional procedure for associating impressions of friction ridge skin by a latent print examiner involves four phases known as Analysis, Comparison, Evaluation, and Verification (ACE-V). This chapter describes the ACE-V process, notes some of its limitations, identifies areas where human factors should be considered, and defines certain terms used throughout this report.

Box 1.1: Terminology

ACE-V: An acronym for Analysis, Comparison, Evaluation, and Verification. The ACE-V process is described in section 1.1.

Bias and *error:* Defined and discussed in section 1.2.

Exemplar or *known prints:* Prints deliberately collected from an individual, usually fingerprints. Exemplar prints can be collected electronically or by using ink on paper cards. Exemplars may be called *ten-prints* when impressions of all ten fingers are taken. Exemplar prints collected during criminal arrests normally include one rolled (from one side of the nail to the other) print of each finger pad and a plain or slap impression of each finger.

Focal point: A small region containing distinguishing features within a print.

Forensic service provider: A laboratory or unit that examines physical evidence in criminal matters and provides testimony and reports about the examination findings. In this report, the term is used interchangeably with *agency*.

Latent print: Unintentional reproduction of the arrangement of ridges on the skin on the underside of the hands or feet made by the transfer of materials from the skin to a surface. This report uses the term *print* or *latent print* to denote impressions from all regions of friction ridge skin unless a more specific term such as “fingerprint” or “palm print” is used.

Latent print examination: The study of latent and exemplar prints to help determine the source of the latent print. Because prints come from the friction ridge area of the skin on the hands or feet, latent print analysis is sometimes referred to as friction ridge analysis. As discussed below, “Analysis” and “Comparison” also have specialized meanings in “ACE-V;” therefore, this report generally uses the term “examination” rather than “analysis” or “comparison” when referring to the totality of work of latent print examiners.

Latent print examiner: The individual who conducts the latent print examination, also called latent print analyst.

Minutiae: Events along a ridge path, including bifurcations (points at which one friction ridge divides into two friction ridges), dots (isolated friction ridge units that have lengths similar their widths), and ridge endings (the abrupt end of ridges), as illustrated in Table 1.1.

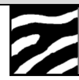
Bifurcation	Dot	Ridge Ending
		

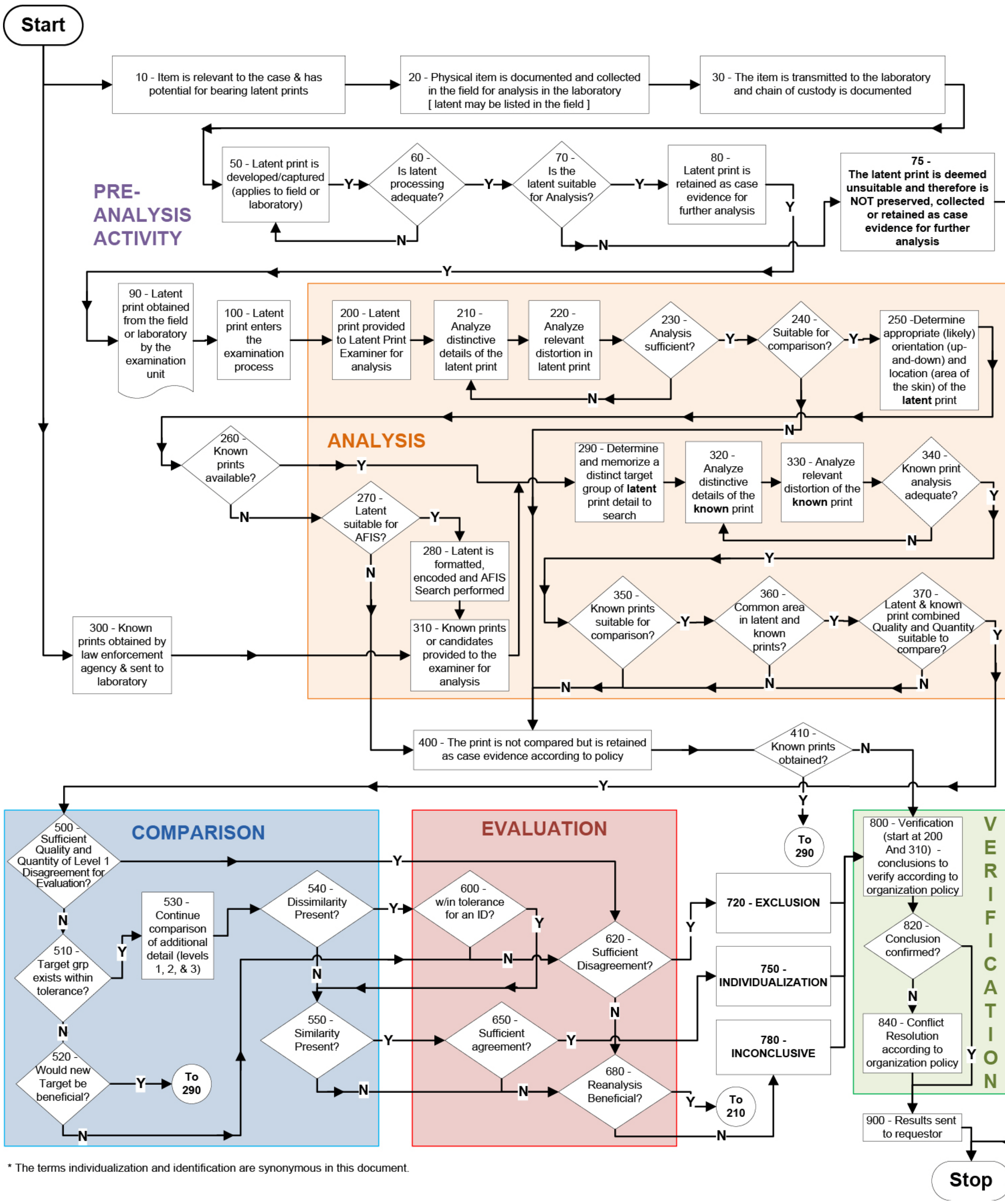
***Table 1.1: Illustrations of some friction ridge minutiae*[8]**

1.1 The ACE-V Process

In broad strokes, a latent print examination using the ACE-V process proceeds as follows: *Analysis* refers to an initial information-gathering phase in which the examiner studies the unknown print to assess the quality and quantity of discriminating detail present. The examiner considers information such as substrate, development method, various levels of ridge detail, and pressure distortions. A separate analysis then occurs with the exemplar print. *Comparison* is the side-by-side observation of the friction ridge detail in the two prints to determine the agreement or disagreement in the details. In the *Evaluation* phase, the examiner assesses the agreement or disagreement of the information observed during Analysis and Comparison and forms a conclusion. *Verification* in some agencies is a review of an examiner's conclusions with knowledge of those conclusions; in other agencies, it is an independent re-examination by a second examiner who does not know the outcome of the first examination.

Figure 1.1, developed by members of the Working Group, describes the steps of the ACE-V process as currently practiced by the latent print examination community. The Latent Print Examination Process Map's purpose is to facilitate discussion about key decision points in the ACE-V process. This chapter briefly describes each step in ACE-V, although the sequence of some of the steps may vary in practice.

[8] Images adapted from Champod, C. *Reconnaissance Automatique et Analyse Statistique des Minuties sur les Empreintes Digitales*. PhD Thesis. Institut de Police Scientifique et de Criminologie, Université de Lausanne, Lausanne, Suisse, 1996.



* The terms individualization and identification are synonymous in this document.

Figure 1.1: The Latent Print Examination Process Map

1.1.1 Analysis

Analysis refers to the initial examination of a friction ridge impression. By inspecting the latent print, the examiner gathers information needed to decide whether it is useful for comparison.

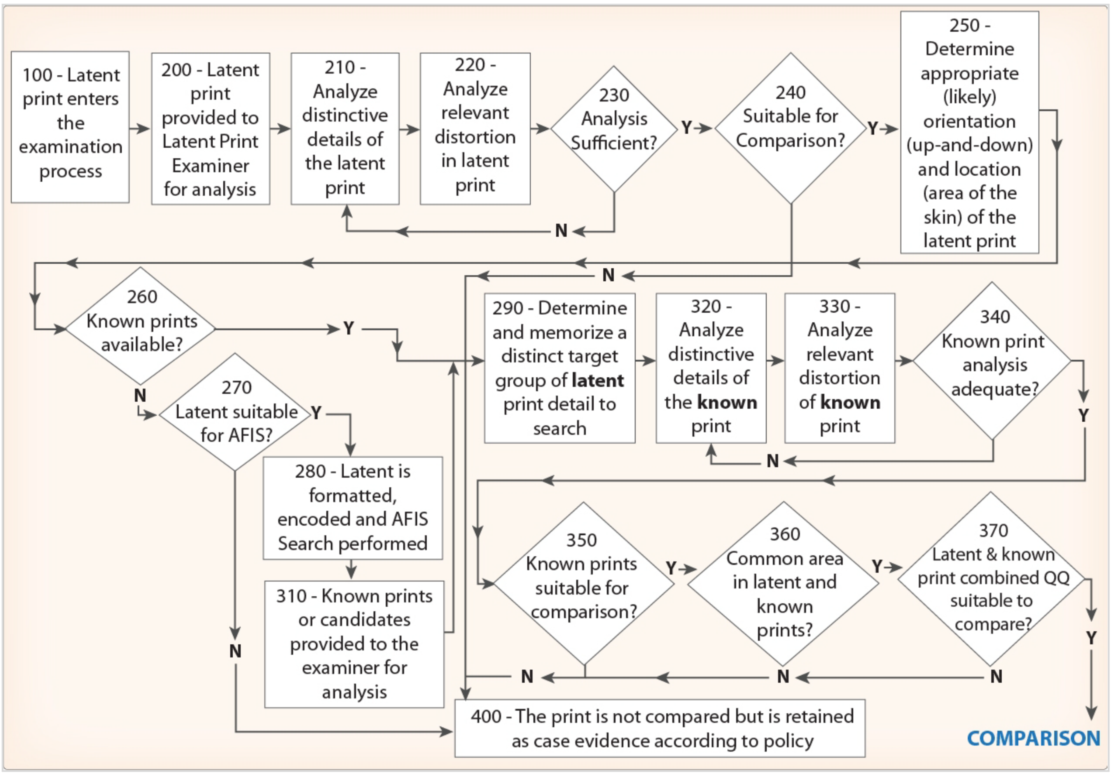


Figure 1.2: Analysis phase of ACE-V

To determine the print's value, the examiner considers three levels of detail in the impression. Level 1 Detail (L1D) is defined as "ridge flow."[9] Ridge flow often translates to a pattern type in a finger or palm, such as a loop, whorl, or arch formation (see Figure 3.1 in Chapter 3); ridge flow also includes other information such as relative curvature. Pattern types are class characteristics shared by many individuals. Level 2 Detail (L2D) is defined as "ridge path."[10] L2D includes, but is not limited to, minutiae, such as ridge endings, bifurcations, or dots. Even the absence of minutiae in an area (called an "open field") can be significant and highly discriminating. Level 3 Detail (L3D) is defined as "ridge shapes." Ridge shapes include the edges of ridges (which may appear indented or protruded) and pores (the location of the center of

[9] Ashbaugh, D. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. CRC Press, 1999.
[10] *Ibid.*

the pore, not the size or shape, which can be highly variable within a source).[11] Chapter 3, which discusses interpreting information in latent prints, provides additional information on the three levels of detail and their use in the Analysis phase.

After considering the details and the distortion, the examiner judges whether the impression is suitable for a comparison. If the examiner concludes that the print lacks sufficient detail for a comparison, then the examination ends with the determination that the latent print is not suitable for a comparison. Otherwise, the examination moves into the Comparison phase.

1.1.2 Comparison

In the Comparison phase, the examiner compares the latent print to one or more exemplar prints. Information gathered in the earlier analysis of the latent print provides a starting point. A comparison of L1D might take only a split second, as when a whorl is present in the latent, but an arch is apparent in the exemplar. If there is no exclusion based upon L1D, then the examiner continues the comparison. If the examiner finds disagreement with respect to the target group that is too extensive to be the result of the distortion noted in the Analysis phase, the examiner will exclude the source of the exemplar as the source of the latent.

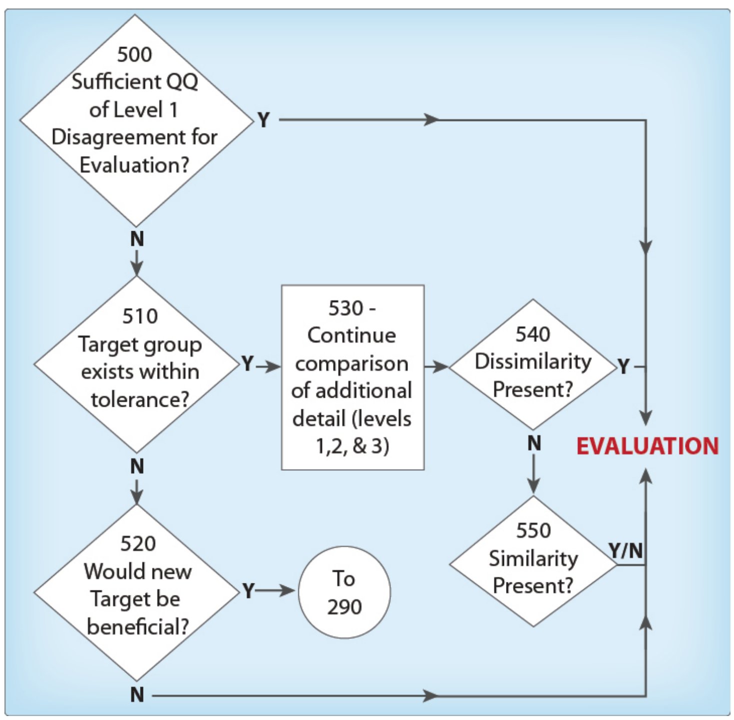


Figure 1.3: Comparison phase of ACE-V

[11] Roddy, A. and J. Stosz. "Fingerprint Features—Statistical Analysis and System Performance Estimates," *Proceedings of the IEEE*, 85, no. 9 (1997): 1390 – 1421; Richmond, S. *Do Fingerprint Ridges and Characteristics Within Ridges Change with Pressure?*. Australian Federal Police, Forensic Services, 2004.

If the initial target group is not found, alternative target groups may be selected. If the examiner locates a comparable set of L1D features in the known exemplar, the examiner proceeds to a detailed, side-by-side comparison of L2D and possibly L3D. If the examiner concludes that the extent of agreement between the two prints satisfies his or her threshold, then the examiner proceeds to the Evaluation phase.

Figure 1.4 displays a latent print (in the middle) and two very similar exemplar prints from monozygotic (identical) twins. One twin is the source of the latent print. These images were used in an unusually difficult inter-laboratory comparison in 1995.

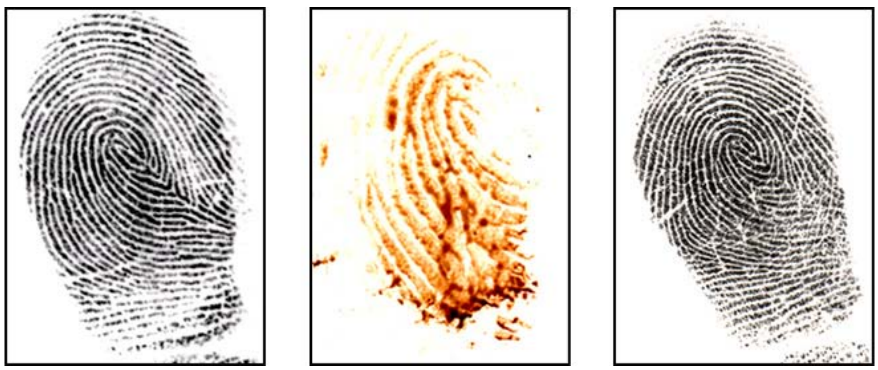


Figure 1.4: A latent print and exemplar prints[12]

[12] Images reproduced and adapted with permission of Collaborative Testing Services.

1.1.3 Evaluation

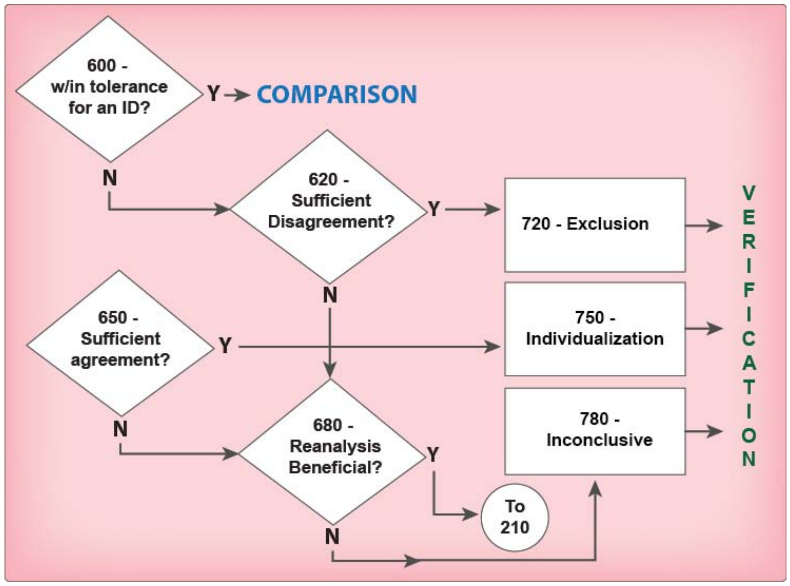


Figure 1.5: Evaluation phase of ACE-V

In the Evaluation phase, the examiner makes the ultimate decision regarding source attribution. Traditionally, three possible findings have been available. First, the examiner can make an "individualization" or "identification." Section 1.2 defines these terms and describes another type of source attribution. Second, the examiner can categorically exclude the latent print by determining that the exemplar print and the latent do not match and cannot share a common source. Finally, the examiner can determine that the information available is inadequate to warrant a conclusion. In that event, the examiner would state that the comparison was "inconclusive" and would provide no additional information about the chances that the two prints share a common source.

The thresholds for these decisions can vary among examiners and among forensic service providers. Some examiners state that they report identification if they find a particular number of relatively rare concurring features, for instance, eight or twelve. Others do not use any fixed numerical standard. Some examiners discount seemingly different details as long as there are enough similarities between the two prints. Other examiners practice the one-dissimilarity rule, excluding a print if a single dissimilarity not attributable to perceptible distortion exists. If the examiner decides that the degree of similarity falls short of satisfying the standard, the examiner can report an inconclusive outcome. If the conclusion is that the degree of similarity satisfies the standard, the examiner reports an identification.

1.1.4 Verification

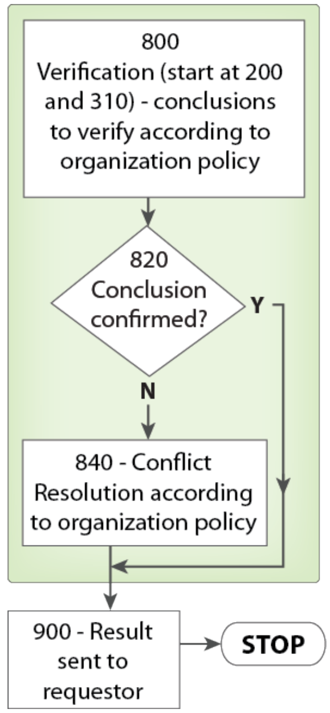


Figure 1.6: Verification phase of ACE-V

In the ACE-V acronym, *V* stands for *Verification*. Verification procedures vary among forensic service providers. At one extreme, the verifier, presented with the first examiner's work, assesses the original conclusion. At the other extreme, the verifier, blinded to the initial examination, performs an independent examination.

1.1.5 Limitations and Concerns about ACE-V

At every step in the ACE-V process, human factors can affect the outcome. Latent print examiners rely heavily on their training and experience to make the required judgments. Subjectivity is an inextricable part of the process. In the Analysis phase, for example, accurate identification of the characteristics that make prints of value depends on the examiner's knowledge, training, and experience. Likewise, in the Comparison phase, variable factors, such as the elasticity of skin and uneven pressure, mean that there will never be perfect congruence between two prints, even if they originate from the same source. The examiner must resolve the question of whether there is sufficient agreement "within tolerance." As Chapter 3 points out, the examiner at least implicitly relies on a sufficiency threshold to resolve that question, and in setting this threshold, the examiner draws on professional knowledge and experience. There is little research at present that provides objective metrics for determining these tolerances.

Of course, the mere existence of subjective elements does not make the process unreliable or invalid. Humans can perform many tasks involving subjective judgments quite accurately and consistently. For example, by holding a heavy book and a much lighter one in each hand, most people can subjectively—but correctly—tell which is heavier. Thus, the mere presence of subjectivity is not a valid criticism of the technique, but it does mean that issues related to human factors can be especially salient to the outcome.

Although ACE-V is a systematic process, meaning that the examination proceeds in an orderly and logical fashion, this does not, by itself, demonstrate that the results are accurate and reproducible. In 2009, a committee of the National Research Council (NRC) stated that ACE-V is "a broadly stated framework for conducting friction ridge analyses. However, this framework is not specific enough to qualify as a validated method for this type of analysis. … Merely following the steps of ACE-V does not imply that one is proceeding in a scientific manner or producing reliable results."[13] Additional study is required to ascertain precisely how well examiners using the process perform under either controlled conditions or in casework, and Chapter 2 describes several possible approaches to developing such information.

Although many in the latent print community describe the ACE-V process as a scientific method (see Chapter 6), the issue is not the label that can or should be attached to the process with respect to human factors. ACE-V is a systematic, skill-based, and widely used process for determining whether two impressions have a common origin. ACE-V designates a logical sequence for a complex process of judgment, but ACE-V itself does not provide substantive guidance about standards to be applied within this sequence. Therefore, even though two examiners might both assert (correctly) that they are using ACE-V, they may be employing different cognitive processes. Those differences create opportunities for human factors to come into play.

1.2 The Meanings of "Bias" and "Error"

The issues of bias and error are critical to assessing the role of human factors in latent print analysis. Those terms are described in detail here, and they are referred to throughout this report.

1.2.1 Bias

The term "bias" has many meanings. This report discusses the term as it is used in three disciplines. In law, "bias" refers to a witness's partiality toward one party (or against another) as a result of financial, emotional, or other interests or attitudes. The law of evidence does not expect all witnesses to be unbiased. Rather, it relies on the disclosure of the biasing interests or attitudes through cross-examination, a procedure that is effective primarily in exposing gross motivational biases (see Chapter 6).

In statistics, "bias" refers to the extent to which an average statistic departs from the parameter it is estimating or to the extent to which measurements on individual units systematically depart

[13] National Academy of Sciences, National Research Council, Committee on Identifying the Needs of the Forensic Science Community. *Strengthening Forensic Science in the United States: A Path Forward*. National Academies Press, 2009, p. 142.

LEXSEE

Copyright (c) 1995 Brooklyn Law School
Brooklyn Law Review

Winter, 1995

61 Brooklyn L. Rev. 1247

LENGTH: 18954 words

ARTICLE: COMING TO GRIPS WITH SCIENTIFIC RESEARCH IN DAUBERT'S "BRAVE NEW WORLD": THE COURTS' NEED TO APPRECIATE THE EVIDENTIARY DIFFERENCES BETWEEN VALIDITY AND PROFICIENCY STUDIES *

Edward J. Imwinkelried **

** Professor of Law, University of California at Davis; former chair, Evidence Section, American Association of Law Schools. Professor Imwinkelried consulted with the plaintiffs in preparing their briefs before the United States Supreme Court in the Daubert case.

SUMMARY:
... For decades, the courts have regarded scientific testimony with suspicion. ... While a validity study tests a scientific technique, in a proficiency study the object of the test is a particular analyst or laboratory. ... Moreover, even in a Frye jurisdiction, the logic of the theory applies only to the passages in the validity study indicating whether or not the expert accepts the theory being tested; the logic does not extend to the passages summarizing the empirical research and quantifying the margin of error. ... In a proficiency study, the researcher is interested in the as-sertive findings by the participating laboratory which are pre-sumably false—the findings at odds with the findings by the referee laboratory. ... When the proponent offers a proficiency study, the theory of logical relevance is that the tendency to err, documented in the study, increases the probability that the analyst erred in ana-lyzing the sample relevant in the pending case; the evidence is relevant to the issue of whether the analyst correctly applied the scientific technique in the instant case. ...

TEXT:
[*1247]

"'If the law supposes that,' said Mr. Bumble . . . 'the law is a ass, a idiot.'"

—Charles Dickens, Oliver Twist

Introduction

For decades, the courts have regarded scientific testimony with suspicion. Their fear was that jurors would uncritically accept such testimony at face value and assign it undue weight. In one case, the California Supreme Court voiced the fear that science is "a veritable sorcerer in our computerized society," a sorcerer who can "cast a spell" over the trier of fact. n1 In another decision, the same court expressed concern about the "misleading aura of certainty which often envelops a new scientific process." n2 For its part, the District of Columbia Court of Appeals asserted that jurors often attribute a "mystic infallibility" to scientific evidence. n3 In a similar vein, the Maryland Court of Appeals has stated that jurors routinely overestimate the probative worth and certainty of scientific testimony. n4 For that matter, several courts have declared that it is doubtful that even legally trained judges are competent to [*1248] pass on essentially scientific questions; in their view, scientists are far better "qualified to assess the general validity of a scientific method." n5

Given these assumptions, it was perhaps to be expected that the courts would turn to the traditional, general acceptance test as the standard determining the admissibility of scientific evidence. n6 Under this standard—sometimes dubbed the Frye test after the 1923 case announcing the standard n7—a scientific theory or technique may not serve as a basis for testimony until the theory or technique has gained general acceptance among the members of the relevant scientific field or specialty. n8 Given courts' fears, the Frye test had two obvious virtues. First, the test is "essentially conservative," n9 striking a note of "judicial caution." Again, the courts assumed that lay jurors ascribe excessive weight to scientific testimony and naively expect it to be virtually infallible. n10 The general-acceptance test helped to ensure that the only testimony admitted would be testimony measuring up to that expectation.

Second, in applying the test, the trial judge does not have to reach the merits of the scientific dispute. Under Frye, judges "could avoid coming to grips with science." n11 The test focuses on an indirect indicator of scientific validity, n12 namely, the popularity of the theory or technique in the specialty field. The existence of a certain degree of popularity is the type of historical, nontechnical issue which judges are accustomed to deciding. General acceptance serves as a surrogate n13 for a direct evaluation of the scientific research underpinning the theory or technique. The test not only permits the trial judge [*1249] to "hide from science;" n14 but it also in effect delegates the decision on admissibility to the scientific community. n15

At one time, the Frye test was controlling in forty-five states and in all the federal circuits. n16 In 1993, however, the United States Supreme Court jettisoned Frye. In that year, the Court handed down its decision in Daubert v. Merrell Dow Pharmaceuticals, Inc. n17 Mr. Justice Blackmun authored the majority opinion. The majority initially ruled that the general-acceptance test is no longer good law in federal court. Justice Blackmun pointed out that under *Federal Rule of Evidence 402*, logically relevant evidence is "admissible, except as otherwise provided by the Constitution of the United States, by Act of Congress, by these rules, or by other rules prescribed by the Supreme Court pursuant to statutory authority." n18 Although the exceptive language in Rule 402 lists the Constitution, statutes, the Federal Evidence Rules, and rules such as the Rules of Civil Procedure "prescribed by statutory authority," Rule 402 makes no mention of case or decisional law. In the past, the Court had held that Rule 402 has the effect of abolishing uncodified exclusionary rules of evidence. n19 Adhering to that holding, Justice Blackmun reasoned that the enactment of the Federal Rules implicitly overturned the Frye test; although the test enjoyed widespread support at common law, the majority stated that there was no language in the text of the Federal Rules which could reasonably bear the interpretation that it codified a general-acceptance standard. The general-acceptance test is a creature of case law.

The majority next ruled that the statutory language of *Federal Rule of Evidence 702* supplies a new, empirical [*1250] validation test to replace Frye. Rule 702 reads, "If scientific . . . knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise." n20 According to Rule 702, the witness's possession of "scientific . . . knowledge" is what qualifies the witness as an expert; and the witness must testify "thereto." Parsing the language of Rule 702, Justice Blackmun concluded that the rule requires the substance of the expert's testimony to qualify as "scientific . . . knowledge."

The question then became defining that term. According to Justice Blackmun, that term does not equate with a particular body of substantive propositions. Rather, it denotes a proposi-tion "derived by the scientific method." n21 The justice described that method as the process of formulating hypotheses and empirically falsifying or validating the hypotheses. n22 A theory or technique constitutes "scientific . . . knowledge" within the intendment of Rule 702 if it rests on sound scientific methodol-ogy. When testimony satisfies this empirical validation test, the testimony is admissible even if the witness's conclusion is novel and controversial. n23 By opening the door to the intro-duction of testimony about novel scientific theories and tech-niques, the United States Supreme Court liberalized the stan-dard for introducing scientific testimony.

In contrast, the California Supreme Court has long been considered a bastion of the traditional, general-acceptance test. The California Supreme Court adopted the test in 1976. n24 The court emphasized that the "conservative nature" of the test was its "primary advantage." n25 The court favored the Frye test because that test "assigned the task of determining reliability of the evolving technique to the members of the scientific com-munity from which the new method emerges." n26 In 1994, the [*1251] court reaffirmed its commitment to the general-acceptance test in People v. Leahy. n27 The court adamantly refused to embrace the Daubert test.

While Leahy ruled that the controlling standard in Califor-nia state court is still the general-acceptance test, the court modified—and arguably liberalized—the test in an important respect. The court acknowledged that some critics of the Frye test have charged that the test reduces the trial judge's deter-mination to a crude "nose count." n28 The court made it clear, however, that it did not want trial judges to administer the test in that fashion. Rather, the court declared that "trial courts, in determining the general acceptance issue, must con-sider the quality, as well as quantity, of the evidence support-ing or opposing a new scientific technique. Mere numerical majority support or opposition by persons minimally qualified to state an authoritative opinion is of little value." n29 The court instructed trial judges to attach greater weight to the views of the "major voices" in the specialty field. n30

Prior to Leahy, several intermediate appellate courts in California had ruled DNA testimony inadmissible. n31 However, seizing upon the language in Leahy, several lower California courts already have ruled that DNA evidence is now admissi-ble in that jurisdiction. n32 (Those rulings might well have influ-enced the Simpson defense team's decision to stipulate initially to the admission of the DNA test results in that case.) As a practical matter, Leahy has relaxed the admissibility standard in California. The lower courts are reading Leahy as a signal that they are free to admit testimony about a scientific theory or technique even in the face of numerically substantial opposi-tion to that theory or technique—so long as the opponents can be characterized as mavericks rather than "major voices" in [*1252] the mainstream of the discipline.

In the words of Judge Alex Kozinski, decisions such as Daubert propel the legal system into a brave new world. n33 Daubert certainly thrusts trial judges into a new world. Justice Blackmun's opinion directs trial judges to eschew surrogates and apply scientific standards in determining the admissibility of proffered expert testimony. n34 In deciding whether to admit the testimony, trial judges must use the same standards scien-tists employ in evaluating the empirical validation of the un-derlying theory or technique. n35 If the expert's hypothesis does not lend itself to empirical testing, a scientist would not accept the hypothesis, and under Daubert, neither should a trial judge permit the introduction of testimony about the hypothesis. n36 Similarly, when the expert has not gone to the length of engag-ing in experimentation or observation to test the hypothesis, a scientist would not regard the hypothesis as validated; and a trial judge should refuse to allow testimony about the hypothe-sis to be submitted to the jury. n37 Shortly after the rendition of the Daubert decision, the Federal Judicial Center released the Reference Manual on Scientific Evidence. n38 The manual's chapters review the rudiments of such scientific specialties as DNA testing, n39 epidemiology, n40 statistical analysis, n41 and toxicology. n42 As the Introduction written by Judge William W. Schwarzer explains, the Center concluded that trial judges needed such a manual because the Daubert "standard demands an understanding by judges of the principles and methods that underlie scientific studies." n43

Like judges, jurors are being thrust into Judge Alex [*1253] Kozinski's brave new world. n44 By liberalizing admissibility standards, decisions such as Daubert and Leahy shift the focus to the question of the weight of the scientific testimony. n45 To attack the weight of scientific testimony, the opponent can proffer studies documenting weaknesses in the underlying technique n46 or the analyst's incompetence in applying the technique. n47 In its highly publicized 1992 report on DNA evi-dence, the National Research Council ("N.R.C.") urged that proficiency studies of DNA laboratories be conducted n48 and that laboratory error rates be disclosed to the jury. n49 Citing the N.R.C. report, the Simpson defense team argued that the jurors in that case should be informed of the studies "estimating frequency of laboratory errors that might cause a false match between samples from different people." n50

It is evident that in the future, scientific research studies can play an important role in both judges' admissibility deci-sions and jurors' determinations of the weight of scientific testimony. However, two other things have become painfully clear in the past few years. First, most judges and attorneys do not appreciate the distinctions among the various types of scientific studies. Second, and worse still, they do not yet real-ize that there are differing evidentiary hurdles to the introduc-tion of the various types of studies.

Numerous types of scientific research studies can become relevant at trial. n51 Unfortunately no universally accepted ter- [*1254] minology describes the various types of studies. n52 Consequent-ly, we must heed Voltaire's imperative and first "define our terms." n53 In particular, we need to define a validity study and a proficiency study for purposes of this Article.

A validity study is designed to measure the accuracy of a scientific technique. n54 The study attempts to identify and quantify the inherent margin of error in the technique; n55 the researcher inquires how often the technique will yield inaccu-rate results even when the analyst strictly follows proper test procedure. By way of example, suppose that the hypothesis is that a new type of breathalyzer validly measures a person's blood alcohol concentration ("BAC"). The researcher could test that hypothesis by using the instrument to gauge the BAC of a number of persons who had consumed alcohol and comparing [*1255] the instrument's readouts with direct blood tests of the same persons. Since direct blood testing has already been estab-lished as a valid method of measuring BAC, n56 a coincidence between the readings and the direct blood test results would tend to verify the hypothesis. However, the readings might disagree with the direct blood alcohol test results to an extent. The degree of disagreement would indicate how frequently the new technique yields inaccurate results n57 even when the ana-lyst follows correct test procedures.

A proficiency study is radically different. While a validity study tests a scientific technique, in a proficiency study the object of the test is a particular analyst or laboratory. The validity test is designed to ensure to the extent possible that correct test procedures are used; in a validity test, the re-searchers attempt to eliminate any concern about the use of a proper test procedure because they want to reach the central question of how often the scientific technique itself will pro-duce inaccurate results despite proper test protocol. In con-trast, a proficiency study endeavors to measure the analyst's proficiency in the sense of the probability that he or she will consistently use proper test procedure. Assuming the validity of the technique, the researcher inquires into the probability that while using the technique, the analyst will "mistakenly use[] the wrong materials or make the wrong measure-ment. . . ." n58 Even when the analyst is utilizing a valid tech-nique, the analyst might commit a "performance-type er-ror[]." n59

The thesis of this Article is that validity and proficiency studies differ both in their scientific nature and their eviden-tiary admissibility. The post-Daubert commentary has general-ly recognized that judges and juries must now learn to come to grips with scientific studies. n60 The commentary has over-looked, however, both the distinction between these types of studies and the evidentiary issues which they pose. The NRC's report urged the admission of testimony about proficiency [*1256] studies without a glimmer of recognition that there might be serious evidentiary hurdles to the admission of the testimo-ny. n61 For that matter, in the Simpson case, testimony about proficiency studies was elicited without objection. The implicit assumption seems to be that these studies are so highly proba-tive in Daubert's brave new world that to paraphrase Mr. Bumble, it would be asinine and idiotic to bar testimony about such studies.

On closer scrutiny, though, it develops that there are sub-stantial, potential evidentiary objections to be made. Part I of this Article is devoted to validity studies. After describing the nature of validity studies, this Part discusses the application of the hearsay rule and the character evidence prohibition to the introduction of testimony about such studies. This Part con-cludes that although the character evidence prohibition should not bar the introduction of the testimony, in many cases the hearsay rule will have precisely that effect. Part II turns to proficiency studies. As in the case of validity studies, this Part initially describes the essential methodology of proficiency studies. Part II then analyzes their admissibility in terms of the same exclusionary rules discussed in Part I, namely, hear-say and character. This Part demonstrates that in an eviden-tiary sense, proficiency studies are the mirror image of validity studies; although the hearsay rule will rarely block the admis-sion of testimony about a proficiency study, in some instances the character-evidence prohibition will prove to be an insupera-ble barrier to admission. Again, there appears to be consensus that judges and juries need the benefit of these studies to cope with scientific evidence in Daubert's brave new world. The Conclusion therefore calls for revising the hearsay and charac-ter doctrines.

I. Validity Studies

A. The Scientific Nature of a Validity Study $TP As noted above, no universally accepted terms of art de-scribe the various types of scientific studies. n62 For that mat- [*1257] ter, a particular scientific research project can be designed to pursue several different objectives. As the terminology is em-ployed in this Article, however, in a validity study, the researchers' objective is to assess the accuracy of a scientific technique. n63 If properly used, does the technique accurately measure "what it is supposed to measure"? n64

What is the logical relevance of a validity study at trial? To answer that question, we must consider the typical struc-ture of a scientist's direct examination. In most cases, the structure is syllogistic. n65 Consider, for example, a case such as the Simpson prosecution in which the government offers DNA evidence. After qualifying an expert in molecular biology, the direct examiner usually organizes the balance of the scientist's testimony along the lines of a syllogism. The expert's major premise would be that when the DNA fragments on two autoradiograms are of the same length and in the same posi-tion, the match indicates that the sources of the two samples share certain DNA markers. n66 The expert's minor premise is the case-specific information. When the scientific testimony is a DNA analysis, the minor premise would be testimony about the two autoradiograms in the case. When the expert applies the major premise to the minor, the result is a conclusion, an opinion relevant to material facts of consequence in the pend-ing case. If the DNA bands on the two autorads matched, the expert would opine that the sources of the two samples had at least those DNA markers in common.

This syllogistic model holds true for "soft" mental-health testimony as well as "hard" instrumental scientific techniques such as DNA analysis. Assume, for instance, that a psychia-trist contemplated testifying that a particular individual was mentally incompetent. After testifying to her credentials, the [*1258] expert would state her major premise. The premise would be a theory about symptomatology: If a person displays symptoms A and B, they suffer from mental illness C. The minor premise would be case-specific information about the symptoms dis-played by the individual in question. The expert might rely on reports from a treating physician and family members. When the expert applies the major premise to the minor, again the application will yield a conclusion or opinion relevant to the competency dispute.

This model enables us to identify the logical relevance of a validity study. The study relates directly to the scientist's ma-jor premise. In the case of DNA analysis, the hypothesis is that when the DNA fragments on two autoradiograms match, the match indicates identical DNA markers. A validity investiga-tion of that hypothesis would attempt to establish the probabil-ity that matching fragments accurately indicate matching DNA markers. In the case of mental-health testimony, the hypothe-sis is that the presence of symptoms A and B in the patient's case history indicates that the patient suffers from mental illness C. In a validity study of that hypothesis, the researcher would endeavor to identify the probability that the concurrence of symptoms A and B accurately predicts the existence of men-tal disorder C.

When the focus is on the expert's major premise, neither the expert's proponent nor the opponent should be limited to validity studies personally conducted by the testifying expert. In this context, the issue is the validity of the theory, not the competence of the testifying expert. Even if the expert is com-petent, the theory or technique can suffer from an inherent margin of error. To assess the validity of the technique used by the testifying expert, it would also be pertinent to consider studies conducted by other scientists. In the scientific tradi-tion, after conducting a validity study, the scientist publishes his or her findings to enable other scientists to replicate the experiment. n67 Indeed, in many cases, the testifying expert per-sonally has not conducted any validity study; rather, he or she [*1259] is simply relying on earlier studies—studies which are perhaps decades old—that validate the hypothesis upon which he or she is relying. The question is the validity of the theory or technique; and whether the study is conducted by the testi-fying expert or a third party, a validity study sheds light on that question. In this respect, a validity study differs funda-mentally from a proficiency study. The whole point of the lat-ter study is to assess the competence of a particular analyst or laboratory; the performance of another analyst or laboratory under even the same conditions does not answer the question of how proficient this analyst or laboratory is.

The two types of studies differ in a further respect. In a validity study, the researcher is comparing the results of prop-erly using the proposed technique with the results yielded by properly using an already validated methodology. Suppose, for example, that the question is the validity of a new drug identi-fication technique. n68 The researcher would employ the new technique to test a number of samples and then compare those test results with results generated by an established methodol-ogy such as gas chromatography/mass spectrometry ("GC/MS"). n69 GC/MS is the "gold standard" in drug identifica-tion technology. n70 In a validity test, the essential terms of the comparison are the results yielded by proper use of the new technique with the results generated by using a proven tech-nique. In contrast, a proficiency test involves a different com-parison. Now the question is the probability that a particular analyst or laboratory uses a particular technique properly. To do so, the researcher compares test results generated by a laboratory correctly using the procedure with results reported by the laboratory whose proficiency is being investigated.

The two types of studies differ in still another respect. If the researcher administers a series of proficiency tests to the same analyst or laboratory to chart their proficiency over time, on each occasion the test conditions should be identical. Other-wise, the researcher is comparing apples and oranges. In con-trast, although all the validity studies testing a particular [*1260] scientific technique will control for the same basic variables, the test conditions can and should vary. More specifically, as the validity studies progress, test conditions should become more and more severe. n71 "The more severe and more diverse the experiments that fail to falsify an . . . hypothesis, the more corroborated . . . it becomes." n72 Hence, a pertinent validity study might not only have been performed by a third party other than the testifying witness; the study might also have been conducted under somewhat different test conditions. n73

B. The Evidentiary Status of a Validity Study

The expert's major premise is obviously an essential com-ponent of his or her reasoning process. If so, the litigants on both sides might have occasion to proffer testimony about a validity study to the judge or jury. Do the technical exclusionary rules of evidence such as hearsay and character apply at this juncture in the trial? If so, can the litigants over-come hearsay and character objections to the introduction of testimony about validity studies?

1. The Applicability of Exclusionary Rules of Evidence

At this point, a reader familiar with the Federal Rules of Evidence might wonder whether this question is much ado about nothing. After all, the last sentence of *Federal Rule of Evidence 104(a)*, governing the trial judge's determination of most foundational or preliminary facts, reads: "In making its determination [the court] is not bound by the rules of evidence except those with respect to privileges." n74 The courts have squarely held that this language renders exclusionary rules such as hearsay inapplicable to foundational testimony. n75

Although *Federal Rule of Evidence 104(a)* does contain that provision, that provision does not reduce this question to a non-issue. To begin with, many jurisdictions do not follow the [*1261] rule stated in the last sentence of Rule 104(a). In these juris-dictions, the exclusionary rules apply even to foundational testimony submitted to the judge when the judge makes ad-missibility determinations outside the jury's presence. n76 Fur-thermore, even when a jurisdiction follows the practice codified in the last sentence of Rule 104(a), that sentence has a limited effect. The sentence authorizes the litigant to submit technical-ly inadmissible information to the judge when the judge is passing on an admissibility question. The sentence does not even purport to authorize exposing the jury to such informa-tion during the trial on the merits. If a litigant wants to pres-ent technically inadmissible information about a validity study to the jury to influence the jury's evaluation of the weight of scientific testimony, the litigant cannot cite to Rule 104(a).

2. Compliance with the Hearsay and Character Exclusionary Rules

Assuming that the litigant must satisfy the technical exclusionary rules, there are two potential objections to the admission of testimony about a validity study: the character evidence prohibition and the hearsay rule.

a. The character evidence prohibition

The character evidence prohibition is codified in *Federal Rules of Evidence 404* and *405*. n77 Rule 404(b) announces that "evidence of . . . acts [other than those alleged in the pleadings] is not admissible to prove the character of a person in order to show that he acted in conformity therewith." n78 The rule prohibits the proponent from introducing evidence of a person's other acts to prove his or her character and, in turn, using the person's character as circumstantial proof of con-duct. n79 The thrust of the prohibition is that the proponent cannot simplistically reason, "He did it once, therefore he did it again." n80 Rule 405(a), however, lifts the character evidence [*1262] ban when the person's character itself becomes "an essential element" in dispute at trial. n81 Suppose that when the propo-nent attempts to present the jury with testimony about a va-lidity study conducted by the testifying witness, the opponent objects on character evidence grounds. The opponent argues that the ultimate issue is whether the expert erred in perform-ing the test conducted in the case and a prior validity study is logically relevant only on a forbidden character-reasoning theo-ry—"He was right (or wrong) before, therefore he was right (or wrong) again."

At first blush, that objection might have some appeal. In the final analysis, however, it is spurious. The character evi-dence prohibition comes into play when a person's character is used as circumstantial proof of the person's conduct on a par-ticular occasion, usually an event such as an accident or crime mentioned in the pleadings. When the proponent proffers a validity study, however, the proponent is not offering the study to prove the analyst's conduct on any occasion. Instead, the focus is the validity of the scientific technique used by the analyst. Even if the analyst dotted every i and crossed every t, the test result might be inaccurate due to the inherent margin of error in the technique; but in that event, the analyst's con-duct is not the cause of the inaccuracy of the analyst's ultimate conclusion.

The situation is akin to cases in which character itself is an essential element. Validity testing is "not a simple pass-fail" n82 or "black and white proposition." n83 The validation pro-cess has "a fundamentally mathematical dimension." n84 Valida-tion is "probabilistic, rather than categorical. The experimental process may disclose a margin of error reflecting the percent-age of cases in which even a qualified, careful analyst [prop-erly] employing the technique will reach an incorrect conclu-sion." n85 Since validation is inherently probabilistic, the trier of [*1263] fact needs to know the probability that the technique will per-form accurately. Ascertaining that probability is the whole point of a validation study.

In this regard, the distinction between a validation and proficiency study is akin to the difference between a negligent manufacture action and a strict product liability case. In the former, the question is whether, in the course of fabricating a particular product, the defendant manufacturer was guilty of negligence. When the plaintiff offers testimony about the defendant's other negligently manufactured goods, the plaintiff is using the other negligent acts to support an inference as to the defendant's conduct on the occasion alleged in the com-plaint. In effect, the plaintiff says, "The defendant did it before, therefore he did it again"—the paradigm of the reasoning for-bidden by the character prohibition. Suppose, however, that the plaintiff offers testimony about similar accidents involving identically designed automobiles or heaters in a strict product liability action. The character prohibition is arguably inappo-site; in this setting,"the focus is on the qualities or properties of a nonperson, usually some kind of physical object. Was a steering wheel's design dangerously defective? To prove the quality of the object, the plaintiff may attempt to introduce evi-dence of other accidents involving the same or similar ob-jects." n86 without running afoul of the spirit of the prohibition. In sum, if the opponent raises a character evidence objection to testimony about a validity study, the trial judge should over-rule the objection.

b. The hearsay rule

Federal Rule of Evidence 802 announces a general rule that hearsay is inadmissible "except as provided by these rules." n87 Rule 801 supplies the definition of hearsay. At the risk of oversimplification, the statutory definition includes as-sertive statements and acts n88 which were made or performed by out-of-court declarants n89 and which are offered at trial to [*1264] prove the truth of the assertion. n90

Assume initially that the jurisdiction in question applies the hearsay rule even to foundational testimony submitted to the judge ruling on the admissibility of proffered scientific evidence. In the face of the hearsay rule, could the judge consider testimony about a validity study which the testifying expert has not personally conducted?

In a Frye jurisdiction such as California, at the admissibil-ity stage the proponent can arguably offer at least part of the study for a nonhearsay purpose. n91 One of the assumptions underlying Frye is that even trial judges are incompetent to pass directly on the question of scientific merit. n92 If the judge finds that the experts disagree sharply over the question of the validity of a scientific theory or technique, the judge is not supposed to determine which position is correct; under Frye, the judge does not resolve the battle of the experts. n93 When the judge finds that there is a substantial controversy, the judge must exclude the testimony; the mere existence of the controversy precludes finding the general acceptance required by Frye. n94 In that light, the proponent can argue that the pas-sages in the study reflecting the expert's acceptance of the theory or technique are logically relevant for a nonhearsay purpose. n95 The mere fact that an expert says he or she sub-scribes to the theory is some evidence of general acceptance, just as their statement that they rejected the theory would be some evidence of controversy.

The best analogy is to the operative fact or verbal act doctrine in contract cases. It is well settled that the statements constituting an offer and acceptance are admissible over a hearsay objection in a contract action. Under the objective theory of mutual assent, it is immaterial whether the defendant meant what he said when he ut-tered the purported offer or acceptance. So long as the defendant's external manifestation of intention matches the plaintiff's, there is a contract. . . . There is a strong parallel between the Frye test and [*1265]

the objective theory of mutual assent. Under the Frye test, the fact that the experts' statements match or do not match is relevant even if the statements are incorrect. If the experts say differing things about the validity of the scientific theory or technique, the difference tends to show a controversy that is the antithesis of general accep-tance. n96

Although this nonhearsay theory enjoys some support in the case law, n97 the theory has limited utility in the present context. To begin with, this theory is unavailable in a Daubert jurisdiction; in such a jurisdiction, general acceptance is no longer in issue even at the admissibility hearing. Moreover, even in a Frye jurisdiction, the logic of the theory applies only to the passages in the validity study indicating whether or not the expert accepts the theory being tested; the logic does not extend to the passages summarizing the empirical research and quantifying the margin of error. Finally, the theory can operate legitimately only at the Frye hearing out of the jury's presence. It is for the judge to decide admissibility under the general acceptance standard; but once the judge rules in favor of admitting the evidence, general acceptance is no longer in issue—the only issue for the jury is the question of whether the theory or technique is in fact valid. If at the trial on the merits the litigant offers an assertive statement that the tech-nique is valid to prove that the technique is valid in fact, the statement is undeniably being used for a hearsay purpose.

In some jurisdictions, a second nonhearsay theory might be tenable at the trial on the merits. Many jurisdictions permit the cross-examiner to confront the expert with passages from texts which appear to contradict the expert's position. n98 These courts permit questioning about contradictory passages at least when the testifying expert consulted the text in the process of forming his or her opinion. n99 While this use of scientific texts is widespread, this theory also has limited utility. The courts sanctioning this practice uniformly state that the cross-examin-er is permitted to resort to the contradictory text in order to [*1266] impeach the testifying expert's credibility. n100 The passage in the text is not admitted as substantive proof of the truth of the assertions in the passage. *California Evidence Code section 721* expressly authorizes cross-examiners to use this impeachment technique, n101 but the official California Revision Commission

Comment to section 721 adds that "the court is required upon request to caution the jury that the statements read are not to be considered evidence of the truth of the propositions stat-ed." n102 Thus, the litigant cannot rely on this theory when he or she wants to use passages in a text documenting a validity study to establish that a particular scientific theory or tech-nique is indeed valid. If the litigant wants to establish the validity of the theory or technique, the litigant must ordinarily turn to the learned-treatise hearsay exception. n103 That is the only exception powerful enough to permit the litigant to bring in passages in an out-of-court validity study as substantive proof on the validity issue. n104 That theory has constrained pa-rameters, however. n105

Federal Rule of Evidence 803(18) represents one of the more liberal versions of the learned-treatise exception. n106 [*1267] Even that version is crabbed, however. By its terms, the rule refers only to "treatises, periodicals, or pamphlets." n107 The common denominator of those three terms is that they all denote published, written material. It would strain that lan-guage beyond the breaking point to extend the statute to justi-fy a witness's reference to another expert's research which had not yet been reduced to writing. Furthermore, the statute ex-pressly states that the material must be "published." n108 Argu-ably, it would not even suffice if the article were in written form but still "in press." Worse still, many states recognize a version of the exception narrower than that set out in Rule 803(18). For example, some jurisdictions limit the scope of the exception to passages in "authoritative" works. n109 Some states are even more restrictive; even if the work itself is a "stan-dard" one, n110 the specific passage in question must state a fact "of general notoriety." n111 As a practical matter, this re-striction limits the exception to judicially noticeable facts. n112 The upshot is that while testimony about a validity study can pass muster under the character evidence prohibition, in many cases the testimony would be excludable as incompetent hear-say. It is true that when litigators adduce testimony about va-lidity studies, more often than not they overlook the hearsay problem. n113 However, if the opponent is acute enough to raise a hearsay objection, and the trial judge applies hearsay doc-trine rigorously, the judge frequently will be obliged to sustain the objection. [*1268]

II. Proficiency Studies

A. The Scientific Nature of a Proficiency Study

At trial, what is the logical relevance of a proficiency study? As Part I explained, the scientist's trial testimony is syllogistic in structure. n114 A validity study bears upon the expert's major premise, that is, the assumption of the sound-ness of the theory or technique upon which the expert relies. A proficiency study relates to a different component of the expert's reasoning process. After describing the theory or tech-nique, the expert specifies the case-specific information to be analyzed and then applies the theory or technique to evaluate that information. At this juncture in the reasoning, the ques-tion becomes how probable it is that the expert properly ap-plied the theory or technique in evaluating the case-specific information. The proficiency study bears on that probability. The theory or technique may yield some inaccurate conclusions even when the analyst correctly applies the technique. A valid-ity study attempts to capture that built-in margin of error. A proficiency study addresses a fundamentally different question, namely, how often will the analyst apply the technique improp-erly.

As we have seen, validity and proficiency studies not only differ in their respective focal point, they also differ in several other respects. In a validity study, the basic comparison is between the results yielded by different scientific techniques. The researcher compares results yielded by the technique be-ing tested with the results generated by a technique that al-ready has been validated. If the researcher is interested in validating a new intoxication testing instrument, he or she could compare its readouts with the results from direct blood alcohol readings. Or if the researcher is investigating the valid-ity of a new drug identification technique, he or she might compare the findings yielded by that technique with a GC/MS analysis. In a proficiency test, the researcher compares the performance of laboratories using exactly the same scientific technique. [*1269]

The Crime Laboratory Proficiency Testing Research Pro-gram, conducted by the Law Enforcement Assistance Adminis-tration ("LEAA") during the 1970s, illustrates the point. n115 Referee laboratories initially analyzed the samples and made certain that they employed proper test procedures. n116 The re-searchers not only instructed the crime laboratories being eval-uated to use the same, standardized procedures as the referee laboratories; n117 they also carefully monitored the manufacture of the samples n118 to ensure that all the samples sent to the participating laboratories were homogeneous. n119 The re-searchers compared the findings by the referee laboratories with the findings reported by the crime laboratories being tested. n120 If the samples were homogeneous, the crime labora-tories were directed to use the very same procedures as the referees; if those laboratories reached different findings than the referee laboratories, the difference would indicate that the laboratories being tested were not properly following the proce-dures. n121 The thrust of this type of inquiry is assessing the competency and performance of the individual laboratory, n122 not an evaluation of the validity of the scientific technique itself.

B. The Evidentiary Status of a Proficiency Study

Like the expert's major premise, the expert's application of that premise to the case-specific information is an integral element of the witness's reasoning process. The parties can have occasion to proffer testimony about a study of the validity of the theory or technique functioning as the major premise. They might also have occasion to proffer testimony about a study of the proficiency of the analyst or laboratory analyzing the case-specific information. Part I demonstrated that at least [*1270] when the parties contemplate presenting the testimony to the jury at the trial on the merits, they must comply with the technical exclusionary rules of evidence. Can the parties intro-duce such testimony over hearsay and character evidence ob-jections?

1. Compliance with the Hearsay Rule

Unlike a validity study, a proficiency study ordinarily will be admissible over a hearsay objection. Consider, for example, a proficiency study conducted by a government agency such as the LEAA's Crime Laboratory Proficiency Testing Program. n123 The report summarizing the test results was prepared by LEAA officials. The report sets out findings by a referee and participating laboratories, compares the findings, and draws conclusions as to the competency of the participating laborato-ries. Upon a moment's reflection, it becomes clear that the admissibility of the LEAA report is a classic problem of double hearsay: the author of the report writes that employees of the referee and participating laboratories made certain assertions to the author. The first level of hearsay is the set of asser-tions by the LEAA officials. n124 That level of hearsay falls squarely within the official-record hearsay exception codified in *Federal Rule of Evidence 803(8)*. n125 In the words of that stat-ute, the LEAA employees' assertions are "matters observed pursuant to duty imposed by law as to which matters there was a duty to report . . . ." n126 Those employees are de jure public officials, and they have firsthand, personal knowledge that the referee and participating laboratories submitted cer-tain findings to them.

Of course, there is a second level of hearsay: the seeming assertions by the employees of the referee and participating laboratories. The admission of the statements by the referee laboratories can again be rationalized under the official-record hearsay exception. The referee laboratories are performing a task for the government agency conducting the proficiency test. [*1271] If the agency had the requisite in-house expertise, the agency could have had its own employees perform the testing; but lacking in-house expertise, the agency delegates the testing to the employees of the referee laboratory. They are consequently acting as de facto public officials, and the official-record hear-say exception extends to statements by de facto as well as de jure officials. n127

Like the assertions by the referee laboratories, the state-ments by the participating laboratories are admissible over a hearsay objection, albeit on a different theory. Here the perti-nent theory is that the statements are being used for a nonhearsay purpose. Under *Federal Rule of Evidence 801(c)*, a statement constitutes hearsay only if its proponent offers it to prove the truth of the assertion contained in the statement. n128 In a proficiency study, the researcher is interested in the as-sertive findings by the participating laboratory which are pre-sumably false—the findings at odds with the findings by the referee laboratory. Those findings indicate the extent of the participating laboratory's incompetence or lack of proficiency. The trier of fact is not being asked to assume that those find-ings by the participating laboratory are true; quite to the con-trary, the trier is invited to assume that those findings are erroneous and then use those erroneous findings to evaluate the participating laboratory's performance level. n129 Given that gauge of the laboratory's capacity, n130 the trier can then make a more informed decision as to whether the laboratory followed proper test protocol in the instant case.

A similar analysis would obtain if the proficiency study were conducted by a private, nongovernmental laboratory. As in the case of a government proficiency test, there are two levels of hearsay, but both levels either fall within a recognized exception or are logically relevant on a nonhearsay theory. n131 In this context, though, the proponent of the proficiency study would rely on the business-entry hearsay exception codified in [*1272] *Federal Rule of Evidence 803(6)*. n132 The first hearsay level is the set of assertions by the employees of the private laboratory conducting the proficiency test. Laboratories certainly qualify as regularly conducted business entities. n133 One of the activi-ties commonly conducted by laboratories is proficiency testing. If the laboratory's employees possess the necessary expertise, they might personally conduct the referee testing.

Assume, though, the more difficult fact situation in which an outside referee laboratory performs that function and both the referee and participating laboratories submit their reports to the employees of the laboratory supervising the proficiency test. It is true that the laboratory's employees have personal knowledge of the contents of the reports submitted to them by the referee and participating laboratories. As in the case of the official-records hearsay exception, however, the reported find-ings by the referee and participating laboratories constitute a second level of hearsay. Can the proponent of the proficiency study surmount a hearsay objection aimed at that level, as he or she can when relying on the official-record exception?

Again, the answer is yes. To begin with, the assertive findings submitted by the referee laboratory will fall within the business-entry exception. If the laboratory conducting the proficiency test hires the referee laboratory, the latter laborato-ry owes the former laboratory a business duty to properly evaluate the samples. "It is unnecessary that the source of the information be a direct employee of the business . . . ." n134 Rather, the test is the existence of a business duty. n135 So long as the person furnishing the information to the business does so pursuant to a business duty, the report is considered to have been generated by the business. Because the referee labo-ratory owes a business duty to the laboratory supervising the proficiency test, the business-entry exception applies to the employees of the referee laboratory. Their reports are therefore admissible as substantive evidence that the reported findings are correct. [*1273]

For their part, the participating laboratories' reports are admissible as nonhearsay. The significant reports are those that disagree with the referee laboratory's findings. The refer-ee laboratory's findings are presumably correct; and to the extent that the participating laboratory's findings disagree, the disagreement reflects adversely on the competency or proficien-cy of the participating laboratory. Its reports are most relevant when they are false and erroneous. In short, the analysis is analogous to the theory used to justify the admission of the results of public-opinion polls. n136 The laboratory is in the business of conducting proficiency tests in the same sense that a pollster is engaged in the business of conducting public-opin-ion surveys; and in both cases, the reports submitted to the business by outsiders—the laboratories being tested or the citizens being polled—can be treated as admissible nonhearsay.

2. Compliance with the Character Evidence Prohibition

In the case of validity studies, use of the potential charac-ter evidence objection is unsound. When we turn to proficiency studies, however, that objection looms much larger. The ulti-mate disposition of the objection turns on whether the proffer of a proficiency study triggers the general rule forbidding char-acter reasoning and, if so, whether the proffer falls within any recognized exception to the general rule.

a. The proffer of a proficiency study as character evidence

Federal Rule of Evidence 404(b) announces a general rule that a litigant may not proffer "evidence of other crimes, wrongs, or acts" by a person "to prove the character of the person in order [in turn] to show [the person's] action in con-formity" with their character. n137 Does the proffer of a profi-ciency study implicate that rule? That query raises three subissues.

First, does a proficiency study amount to "evidence of [*1274] other crimes, wrongs, or acts" within the meaning of that ex-pression in Rule 404(b)? Two leading commentators have ar-gued that as a matter of evidentiary policy, the scope of the character evidence prohibition should be limited to "morally tinged" conduct. n138 Proficiency studies document "perfor-mance-type errors." n139 Such errors, however, are hardly crimes; and after all, "to err is human[]." n140 Does evidence of such errors constitute proof of a "crime[], wrong[], or act[]"? That question must be answered in the affirmative.

The testimony constitutes "evidence of a wrong[]." In this context, an error can amount to an actionable civil wrong. The courts routinely entertain tort actions based on errors in laboratory analysis. For example, a test subject can maintain a negligence action against a drug testing laboratory which erro-neously reports to the subject's employer that the subject uses illegal drugs. n141 Libel law points to the same conclusion. It is not only libelous to assert that a businessperson is incompe-tent; it is libelous per se, that is, actionable without proof of special damage. n142

Furthermore, as a matter of statutory construction, it is difficult to embrace the argument that the character evidence prohibition should be limited to "morally tinged" conduct. n143 As a matter of policy, it makes sense to confine the prohibition to such conduct. One rationale for the prohibition is that the routine admission of evidence of a person's bad character would prejudice the trier of fact and tempt the trier to decide the case against the person in order to punish them for past misdeeds. n144 The risk of prejudice is minimal if the conduct in question is neither criminal nor tortious. Yet, Rule 404(b) ex- [*1275] pressly refers to "crimes, wrongs, or acts" in the alterna-tive. n145 It is a well-settled maxim of statutory interpretation that courts should prefer a construction that gives effect to every word in a statute. n146 The courts eschew interpretations that render a term inoperative n147 or superfluous. n148 Judicial redaction of a statute is an extraordinary step n149 which a court should take only to avoid a truly absurd result or the frustration of a clearly expressed legislative intent. n150 As a matter of evidentiary policy, it would probably be defensible to restrict the character evidence prohibition to testimony about other crimes or civil wrongs. However, the question is not simply one of common-law policy. Rather, the issue is a ques-tion of statutory construction, and Congress chose to insert the word "acts" in Rule 404(b) in addition to "crimes" and "wrongs."

Since the proffer of a proficiency study does amount to evidence of "wrongs, or acts," we must reach the second sub-issue. That question is whether, in the words of Rule 404(b), the evidence is being used "to prove the character of a per-son." n151 Like "wrongs or acts," the term "person" poses a definitional problem. Does "person" apply only to natural per-sons such as a particular laboratory technician, or does the term extend to entities such as the laboratory itself?

Any court would agree that the litigant is offering testimo-ny about the conduct "of a person" if the litigant attempted to [*1276] introduce evidence of a particular technician's errors on a prior proficiency study to increase the probability that the same technician erred in analyzing the samples in the instant case. The more troublesome issue is whether the prohibition is ap-plicable when the litigant proffers a proficiency study and it is unclear whether the analyst who performed the test in the case at bar participated in the proficiency study. It is unsettled whether the term "person" includes entities such as incorporat-ed laboratories. n152 In popular usage, we rarely allude to the "character" of an entity. n153 However, Rule 404(b) was modeled after *California Evidence Code section 1101*, the governing statute in the Simpson case. n154 Like Rule 404(b), section 1101 uses the term "person." In a separate section, however, the California Evidence Code sets out a definition of "person;" and that definition explicitly includes a "firm, association, organiza-tion, partnership, business trust, corporation, limited liability company, or public entity." n155

In the present setting, it would be especially wrong-mind-ed to refuse to include a laboratory within the protection of the character evidence prohibition. As previously stated, it is clear that the prohibition applies when the litigant proffers a partic-ular analyst's prior errors to increase the probability that the analyst erred again. If the prohibition applies and there is no exception, the prohibition renders the evidence inadmissible. Thus, the outcome is that the evidence would be barred even though it tended to show the analyst's personal incompetence.

Contrast the outcome on the assumption that a laboratory is not a "person" covered by Rule 404(b). On that assumption, the prohibition is inapplicable; and the evidence is admissible. The evidence would be admitted even though it has much less probative value than evidence of the analyst's personal incom-petence. When the evidence takes the form of the laboratory's proficiency test, in the final analysis the evidence is admitted to show incompetence by association; even though the laboratory's analyst who conducted the test in this case might not have participated in the laboratory's earlier proficiency [*1277] study, the study's results would be admitted to attack the competence of his or her analysis in this case. Hence, excluding a laboratory from the meaning of "person" leads to an absurd result: the prohibition excludes highly probative evidence of a particular analyst's personal incompetence but permits the admission of evidence that is relevant only on an incompe-tence-by-association theory.

If the litigant offered evidence of the analyst's personal incompetence or the court opted to extend the prohibition to entities such as laboratories, we would reach the third and last subissue, namely, whether the evidence is being utilized "to show [the person's] action in conformity" with character. n156 As previously stated, the prohibition forbids litigants from using a person's character as circumstantial proof of conduct. n157

Can the litigant argue that this is one of the extraordinary situations in which character itself is in issue? *Federal Rule of Evidence 405(b)* lifts the bar of the character evidence rule in the rare case "in which character or a trait of character of a person is an essential element of a charge, claim, or de-fense." n158 That question must be answered in the negative. When the proponent offers a proficiency study, the theory of logical relevance is that the tendency to err, documented in the study, increases the probability that the analyst erred in ana-lyzing the sample relevant in the pending case; the evidence is relevant to the issue of whether the analyst correctly applied the scientific technique in the instant case. Assume that the laboratory has misanalyzed other samples in other cases. Nev-ertheless, the trier of fact should accept the laboratory's find-ings if the trier concludes that the analyst followed proper test protocol in the instant case.

This is not a case in which the pleadings or substantive law place the laboratory's competence directly in issue. If a newspaper published an article assailing the laboratory's profi-ciency, the laboratory might sue for libel, n159 and the newspa-per could defend on the ground that its report was true. n160 Given those pleadings, the laboratory's competence itself would [*1278] be one of the facts in issue. *Federal Rule of Evidence 405* would come into play, n161 and the trial judge could undoubted-ly overrule a character evidence objection to the proffer of a proficiency study.

This is the juncture in the analysis at which the distinc-tion between proficiency and validity studies emerges most starkly. Validation is an intrinsically probabilistic process; and once the validity of a scientific technique comes into issue, the trier needs to know the probability that the technique will yield an inaccurate result even if the laboratory meticulously complies with proper test procedure. When the trier is evaluat-ing the validity of the scientific technique itself, the trier need not draw any inference as to the analyst's conduct; to the ex-tent of its invalidity, the technique can produce an erroneous result even when the analyst's conduct is flawless.

The logical relevance of a proficiency study is fundamen-tally different. When the proponent offers a proficiency study, the proponent is attempting to show initially that the laborato-ry is prone to error and ultimately that the laboratory once again committed a "performance-type error[]." n162 Beyond any cavil, the proponent is inviting the trier of fact to draw the intermediate inference that the analyst has a character trait, disposition or propensity for such errors. This is quintessential character reasoning. The character prohibition forecloses "once a thief, always a thief" reasoning, n163 and in principle it should preclude a litigant from arguing "once an incompetent, always an incompetent."

b. The applicability of an exception to the character evidence prohibition

If the proffer of a proficiency study would otherwise violate the character prohibition, the proponent's only hope for defeat-ing a character objection is convincing the judge that there is a pertinent exception to the character ban. After announcing the [*1279] general character ban, *Federal Rule of Evidence 404(a)* recog-nizes several exceptions to it. n164 The first two exceptions are inapposite here; they relate to the character of a criminal ac-cused or a named victim in certain types of criminal cases. n165 The third exception, though, is of interest. Rule 404(a)(3) codi-fies that exception, allowing "evidence of the character of a witness, as provided in rules 607, 608, and 609." n166 The pro-ponent of a proficiency study might argue that this exception justifies receipt of testimony about the study; the thrust of the argument would be that this exception permits testimony about character on a credibility theory of logical relevance and that the study is relevant on that very theory.

Unfortunately, *Federal Rule of Evidence 607* sheds little light. That statute addresses the question of who may impeach a witness, n167 but it says nothing about how a witness may be impeached. Rules 608 and 609, however, provide insight into the latter question. Rule 608(a) expressly authorizes the re-ceipt of "opinion and reputation evidence of character." n168 Significantly, though, Rule 608(a) specifies that "the evidence may refer only to the character trait for truthfulness or untruthfulness." n169 Rule 608(b) permits cross-examination about certain types of specific acts even if they have not yet re-sulted in a conviction. n170 In an important respect, however, Rule 608(b) parallels Rule 608(a); Rule 608(b) states that the acts in question must be "probative of truthfulness or untruth-fulness." n171 Rule 609 governs when the impeaching evidence takes the form of the witness's prior conviction for a crime. n172 Rule 609(a)(1) permits inquiry about felony convictions when the judge concludes that the probative value of the conviction evidence outweighs any attendant probative dangers. n173 Last-ly, Rule 609(a)(2) sanctions "evidence that any witness has [*1280] been convicted of a crime . . . involving dishonesty or false statement." n174

These provisions share two common denominators. Admit-tedly, one is that they authorize the receipt of testimony logi-cally relevant to a witness's credibility rather than the histori-cal merits of the case. The second common denominator is that the provisions focus specifically on the character trait of un-truthfulness. Rules 608(a) and 608(b) refer expressly to that character trait. n175 Rule 609(a)(2) uses even narrower lan-guage, limiting its scope to offenses "involving dishonesty or false statement." n176 For that matter, in the final analysis, even Rule 609(a)(1) targets that character trait. The theory of logical relevance underlying felony impeachment under 609(a)(1) is that the conviction demonstrates the witness's willingness to violate important social norms and that that willingness increases the probability that the witness would "again violate a[n important] social norm and testify untruth-fully." n177 The Advisory Committee Note to Rule 609 explains that the committee drafted 609(a)(1) on the assumption that a "demonstrated instance of willingness to engage in conduct in disregard of accepted patterns . . . translates into willingness to give false testimony." n178 Further, in striking the balance between the conviction's probative worth and the incidental probative dangers, the trial judge should give particular atten-tion to the question of whether the nature of the felony bears directly on the character trait of truthfulness. n179

In this light, it is misleading to suggest that Rule 405(a)(3) broadly authorizes the receipt of character evidence relevant on a credibility theory. Quite to the contrary, Rule 405(a)(3) represents a narrow exception precisely targeting evidence logically relevant to a witness's character trait of untruthful- [*1281] ness. The ordinary meaning of "untruthfulness" is a disposition to consciously lie. n180 There are cases in which scientific wit-nesses testify untruthfully, for example, by exaggerating cre-dentials. n181 However, in most cases—particularly in the cases identified in proficiency studies—the cause of the error is a blunder pure and simple. The errors shown by proficiency studies relate to a witness's character trait for competence, but they do not pertain to the character trait for untruthful-ness—the focal point of the limited exception codified in Rule 405(a)(3).

The drafters of the Federal Rules used the California Evi-dence Code as one of their principal models. n182 The result is even clearer under that statutory scheme. For example, Evi-dence Code section 786 proclaims: "Evidence of traits of his character other than honesty or veracity, or their opposites, is inadmissible to attack or support the credibility of a wit-ness." n183 Another provision, Evidence Code section 1104, ex-pressly extends the character prohibition to "evidence of a trait of a person's character with respect to care or skill. . . ." n184 The California Law Revision Commission Comment to section 1104 asserts that by virtue of that provision, "character evi-dence with respect to care or skill is inadmissible to prove that conduct on a specific occasion was either careless . . . or un-skilled. . . ." n185 The Comment makes it clear that the judge has no discretion to admit evidence probative only of that prop-osition; the Comment states flatly that section 1104 prescribes "a fixed exclusionary rule." n186 Of course, as we have seen, a proficiency study is logically relevant only on that theory; and the plain statutory mandate is therefore that judges reject proffered testimony about such studies. [*1282]

Conclusion

It has been observed that evidence law should be struc-tured to ensure that experts educate the trier of fact. n187 That model certainly seems preferable to a system under which experts merely express opinions ipse dixit and expect the trier to uncritically defer to the opinion as a matter of course. n188 If the former model is preferable, it seems to follow as a corollary that testimony about scientific validity and proficiency studies should be admissible. If the jury is to decide independently whether to conclude that a scientific technique is valid, the jury needs to know the probability that the technique will yield inaccurate results. Likewise, if the jury must determine wheth-er a laboratory correctly applied a scientific technique, the jury needs a sense of the laboratory staff's proficiency in using the technique. On that issue, the most trustworthy evidence would be a proficiency study investigating the staff's competency.

The case for admitting testimony about validity and profi-ciency studies is compelling if we accept the widespread prem-ise that, by and large, laypersons tend to overestimate the probative value of scientific testimony. Many lower courts subscribe to that premise. n189 In Daubert, n190 the Supreme Court approvingly quoted Judge Jack B. Weinstein's statement that expert testimony can be "quite misleading because of the difficulty in evaluating" its probative worth. n191 If that fear is well-founded, it would seem imperative to ensure the admissi-bility of validity and proficiency studies. Evidence that the technique sometimes produces erroneous results would coun-teract the risk that the jurors would assume that the scientific technique is infallible. n192 Likewise, testimony about a profi- [*1283] ciency study would be an effective antidote for the facile as-sumption that a laboratory's staff has absolutely mastered the protocol for using a technique.

As previously stated, in its 1992 report the N.R.C. simply assumed that evidence law permits the introduction of testimo-ny about proficiency studies. n193 In the Simpson case, the tes-timony was admitted without objection. Like Mr. Bumble, the bench and litigation bar apparently cannot believe that evi-dence law would be so asinine as to bar evidence of such sci-entific studies. There are modern day Bumbles. In a 1983 deci-sion, the Criminal Division of the English Court of Appeal faced an evidentiary objection to testimony about a scientific study conducted by the Home Office. n194 The court overruled the objection. As one commentator remarked, it seemed patent to the court "as a matter of common sense" that the study had to be admitted. n195 In her words, "a legal system that would . . . rule out such cogent evidence . . . [could not be] defended." n196 As we have seen, though, if the exclusionary rules of evidence are rigorously applied to testimony about validity and proficiency studies, common sense might not pre-vail; there are potentially successful hearsay objections to evidence of validity studies, and the character evidence prohi-bition might block the admission of evidence of a relevant proficiency study. If we are to ensure the admission of these seemingly necessary types of evidence, the hearsay and charac-ter rules should be revised accordingly.

It is understandable that neither the bench nor the bar has yet recognized the evidentiary hurdles to the introduction of testimony about scientific research studies. Again, one of the beauties of the old Frye regime was that under that evidentia-ry standard, judges and litigators could "hide from sci-ence." n197 The general-acceptance standard enabled judges and litigators to "avoid coming to grips with science." n198 Daubert has ushered in a new era in which the legal system no longer has that luxury. Judges and litigators will need some time to [*1284] get their bearings; they need to become accustomed to working with scientific studies. As Judge Alex Kozinski noted, judges and litigators face the "daunting task" n199 of evaluating scien-tific research n200 in the post-Daubert Brave New World. n201

It was, of course, the 20th century English novelist, Aldous Huxley, who first gave us a vision of Brave New World. n202 Perhaps, though, the most telling remark was made by his forebear, the famous 19th century biologist, Thomas Huxley. The latter once wrote that "the rung of a ladder was never meant to rest upon, but only to hold a man's foot long enough to enable him to put the other somewhat higher." n203

Daubert was a step in the right direction, at long last forcing the courts to confront directly the scientific standards for determining the merit of expert testimony. n204 It is now time for the next step—another step upward toward a more so-phisticated appreciation of the critical scientific and evidentia-ry differences between validity and proficiency studies.

FOOTNOTES:

n1 *People v. Collins, 438 P.2d 33, 33 (Cal. 1968).*

n2 *People v. Kelly, 549 P.2d 1240, 1245 (Cal. 1976).*

n3 *United States v. Addison, 498 F.2d 741, 744 (D.C.Cir. 1974).*

n4 *Reed v. State, 391 A.2d 364, 370 (Md. 1978).*

n5 E.g., *Kelly, 549 P.2d at 1244-45.*

n6 1 Paul C. Giannelli & Edward J. Imwinkelried, Scientific Evidence sections 1-5 (2d ed. 1993).

n7 Id.

n8 Id.

n9 *People v. Kelly, 549 P.2d 1240, 1244-45 (Cal. 1976).*

n10 Edward J. Imwinkelried, The Standard for Admitting Scientific Evidence: A Critique from the Perspective of Juror Psychology, *28 Vill. L. Rev. 554, 562 (1983).*

n11 Bert Black et al., Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge, *72 Tex. L.Rev. 715, 718 (1994).*

n12 Ronald L. Carlson et al., Evidence in the Nineties: Cases, Materials and Problems 289 (3d ed. 1991).

n13 Black et al., supra note 11, at 721, 725.

n14 Black et al., supra note 11, at 722.

n15 *People v. Kelly, 549 P.2d 1240, 1244-45 (Cal. 1976).*

n16 Betty R. Steingass, Changing the Standard for the Admissibility of Novel/Scientific Evidence: State v. Williams, *40 Ohio St. L.J. 757, 769 (1979).*

n17 *113 S. Ct. 2786 (1993).*

n18 *Fed. R. Evid. 402.*

n19 See generally Edward R. Becker & Aviva Orenstein, The Federal Rules of Evidence After Sixteen Years--The Effect of "Plain Meaning" Jurisprudence, the Need for an Advisory Committee on the Rules of Evidence, and Suggestions for Selective Revision of the Rules, *60 Geo. Wash. L. Rev. 857 (1992);* Randolph N. Jonakait, The Supreme Court, Plain Meaning, and the Changed Rules of Evidence, *68 Tex. L. Rev. 745 (1990);* Glen Weissenberger, The Supreme Court and the Interpretation of the Federal Rules of Evidence, *53 Ohio St. L.J. 1307 (1992)*.

n20 *Fed. R. Evid. 702*.

n21 *Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2795 (1993)*.

n22 *Id. at 2795-96*.

n23 Edward J. Imwinkelried, The Daubert Decision: Frye Is Dead, Long Live the *Federal Rules of Evidence, 29* Trial 60, 64 (Sept. 1993).

n24 *Kelly, 549 P.2d 1240*.

n25 *Id. at 1244*.

n26 Id.

n27 *882 P.2d 321 (Cal. 1994)*.

n28 *Id. at 330*.

n29 *Id. at 336-37*.

n30 Id.

n31 E.g., *People v. Barney, 10 Cal. Rptr. 2d 731 (Cal. Ct. App. 1992)*.

n32 *People v. Taylor, 40 Cal. Rptr. 2d 132 (Cal. Ct. App. 1995); People v. Marlow, 39 Cal. App. 4th 343, 41 Cal. Rptr. 2d 5* (Cal. Ct. App.), superseded by grant of review, *899 P.2d 65 (Cal. 1995); People v. Wilds, 37 Cal. Rptr. 2d 351* (Cal. Ct. App.), superseded by grant of review, *890 P.2d 1115 (Cal. 1995); People v. Soto, 35 Cal. Rptr. 2d 846 (Cal. Ct. App. 1994),* superseded by grant of review, *890 P.2d 1115 (Cal. 1995)*.

n33 *Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F.3d 1311, 1315* (9th Cir.), cert. denied, *116 S. Ct. 189 (1995)*.

n34 Bert Black, The Supreme Court's View of Science: Has Daubert Exorcised the Certainty Demon?, *15 Cardozo L. Rev. 2129 (1994)*.

n35 Black, supra note 34, at 2131.

n36 *Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2796-97 (1993)*.

n37 Id.

n38 Fed. Jud. Ctr., Reference Manual on Scientific Evidence (1994) (herein-after "Reference Manual").

n39 Id. at 273.

n40 Id. at 121.

n41 Id. at 331.

n42 Id. at 181.

n43 Reference Manual, supra note 38, at 2.

n44 *Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F.3d 1311, 1315* (9th Cir.), cert. denied, *116 S. Ct. 189 (1995)*.

n45 Edward J. Imwinkelried, A New Era in the Evolution of Scientific Evi-dence—A Primer on Evaluating the Weight of Scientific Evidence, *23 Wm. & Mary L. Rev. 261 (1981)*.

n46 Edward J. Imwinkelried, The Methods of Attacking Scientific Evi-dence section 10-6 (2d ed. 1992).

n47 Id. at Ch. 12.

n48 Comm. on DNA Technology in Forensic Science, Nt'l Research Coun-cil, DNA Technology in Forensic Science S-14, S-15, 2-5, 3-23 (1992).

n49 Id. at 3-17, 3-25.

n50 Notice of Objections to Testimony Concerning DNA Evidence and Memoran-dum in *Support Thereof, People v. Simpson, Case No. BA 097211, 1995 WL 126286* at *9 (Super. Ct., L.A. County, March 20, 1995).

n51 The terminology for describing the various types of studies is not particular-ly well settled. See Bert Black, A Unified Theory of Scientific Evidence, *56 Fordham L. Rev. 595, 599, 612 (1988)*. However, one leading commentator, Profes-sor Paul C. Giannelli, states:

Although courts use the term "validity" and "reliability" interchangeably, the terms have distinct meanings in scientific jargon. "Validity" refers to the ability of a test procedure to measure what it is supposed to mea-sure—its accuracy. "Reliability" refers to whether the same results are obtained in each instance in which the test is performed—its consistency. Validity includes reliability, but the converse is not necessarily true.

Paul C. Giannelli, The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later, *80 Colum. L. Rev. 1197, 1201 n. 20 (1980)*. As Pro-fessor Giannelli uses the term "reliability," a reliability study is designed to assess consistency. If the objective were to measure a particular analyst's consistency, the study would require the analyst to repeat essentially identical forensic tasks, and the researcher would gauge the probability that the analyst would reach the same findings. If the objective were to measure interanalyst consistency, the study would require different analysts to perform the same forensic tasks, and the researcher would measure the probability that the various analysts would reach the same findings. An interanalyst reliability study could be highly relevant in a professional negligence action. Assume, for instance, that the plaintiff filed an action against a laboratory for alleged negligence in conducting a test. The plaintiff introduces the testimony of an analyst from another laboratory who retested the sample and ar-rived at diametrically opposed findings. The defending laboratory might even con-cede that its test outcome was in error but argue that they were nevertheless not guilty of negligence. To support their argument, they might proffer an interanalyst reliability study indicating that even the state-of-the-art scientific technique which they used can produce inconsistent results in the hands of equally well-qualified, meticulous analysts.

n52 Black, supra note 51, at 599, 612; James E. Starrs, Frye v. United States Restructured and Revitalized: A Proposal to Amend *Federal Evidence Rule 702*, *26* Jurimetrics 249, 255-58 (1986).

n53 Quoted in The Home Book of Quotations: Classical and Modern 428 (10th ed. 1967).

n54 Giannelli, supra note 51.

n55 Edward J. Imwinkelried, Judge Versus Jury: Who Should Decide Questions of Preliminary Facts Conditioning the Admissibility of Scientific Evidence?, *25 Wm. & Mary L. Rev. 577, 601 (1984)*.

n56 2 Paul C. Giannelli & Edward J. Imwinkelried, Scientific Evidence, at section 22-3(B), at 205-07 (2d ed. 1993).

n57 Black, supra note 51, at 637-38.

n58 Black, et al., supra note 11, at 775.

n59 Black, et al., supra note 11, at 775.

n60 See supra note 44 and accompanying text.

n61 See supra notes 48-49 and accompanying text.

n62 See supra note 52 and accompanying text.

n63 Giannelli, supra note 51. See also Annot. Daniel A. Klein, Reliability of Sci-entific Technique and Its Acceptance Within Scientific Community, as Affecting Ad-missibility, at Federal Trial, of Expert Testimony as to Result of Test or Study Based on Such Technique—Modern Cases, *105 A.L.R. Fed. 299, 319 (1991)* ("ex-periments to verify the accuracy of the technique").

n64 Giannelli, supra note 51.

n65 Edward J. Imwinkelried, The Educational Significance of the Syllogistic Structure of Expert Testimony, *87 Nw. U. L. Rev. 1148 (1993);* Edward J. Imwinkelried, The "Bases" of Expert Testimony: The Syllogistic Structure of Scien-tific Testimony, *67 N.C. L. Rev. 1 (1988)*.

n66 2 Giannelli & Imwinkelried, supra note 56, at 1-39.

n67 Ernest E. Snyder, History of the Physical Sciences 38-39 (1969); Andre A. Moenssens, Admissibility of Scientific Evidence—an Alternative to the Frye Rule, *25 Wm. & Mary L. Rev. 545, 556 (1984);* James W. Osterburg, The Scientific Method and Criminal Investigation, 9 J. Pol. Sci. Admin. 135 (1981).

n68 2 Giannelli & Imwinkelried, supra note 56, at 271-339.

n69 2 Giannelli & Imwinkelried, supra note 56, at 313-19.

n70 David J. Greenblatt, Urine Drug Testing: What Does it Test?, *23 New Eng. L. Rev. 651, 655* (1988-89).

n71 Black et al., supra note 11, at 762-63.

n72 Black et al., supra note 11, at 783.

n73 Black et al., supra note 11, at 762-63, 784.

n74 *Fed. R. Evid. 104*.

n75 *State v. Cardone, 368 A.2d 952, 954-55 (N.J. Super. Ct. App. Div. 1976),* cert. denied, *379 A.2d 234 (N.J. 1977)*.

n76 Edward J. Imwinkelried et. al. California Evidentiary Foundations 4 (2d ed. 1994).

n77 *Fed. R. Evid. 404 - 05*.

n78 *Fed. R. Evid. 404*.

n79 Carlson, supra note 13, at 451-52.

n80 Edward J. Imwinkelried, Uncharged Misconduct Evidence sections 1:03, 2:18 (1984).

n81 *Fed. R. Evid. 405*.

n82 Black, et al., supra note 11, at 762.

n83 Black, et al., supra note 11, at 785.

n84 B.J. George, Jr., Statistical Problems Relating to Scientific Evidence in Sci-entific and Expert Evidence 105, 118 (Edward J. Imwinkelried ed., 2d ed. 1981).

n85 Imwinkelried, supra note 55, at 602.

n86 Imwinkelried, et al., supra note 76, at 200.

n87 *Fed. R. Evid. 802*.

n88 *Fed. R. Evid. 801(a)*.

n89 *Fed. R. Evid. 801(b)*.

n90 *Fed. R. Evid. 801(c)*.

n91 Edward J. Imwinkelried, Forensic Science: Role of the Hearsay Doctrine in Litigating Frye Challenges to the Admissibility of Scientific Evidence, *29 Crim. L. Bull. 158* (Mar.-Apr. 1993).

n92 *Id. at 162-63*.

n93 *Id. at 163*.

n94 Id.

n95 *State v. Cary, 239 A.2d 680, 684 (N.J. Super. Ct. Law Div. 1968)*.

n96 Imwinkelried, supra note 91, at 163.

n97 *Cary, 239 A.2d at 684*.

n98 Edward J. Imwinkelried, The Methods of Attacking Scientific Evi-dence section 4-4(B), at 95-98 (2d ed. 1992).

n99 Id.

n100 Id. at 95.

n101 *Cal. Evid. Code section 721*.

n102 Parker's Evidence Code of California 88 (1985).

n103 The business-entry exception, codified in *Federal Rule of Evidence 803(6)*, is usually inapplicable. That exception comes into play when a business's employee gains personal knowledge of a fact and, in the regular course of business, prepares a record documenting that fact. With some exceptions, the person with first-hand knowledge of the fact must be an employee of the business. See generally Edward J. Imwinkelried, et. al., Courtroom Criminal Evidence section 1220 (2d ed. 1993). On rare occasions, a validity study will fall within the ambit of this exception. Suppose, for example, that the researcher is on the faculty of the University of California and the University of California Press publishes the report detailing the study. In most cases, the researcher with personal knowledge of the study has no formal business relationship with the entity publishing the study. If the study ap-pears in book form, the publisher might be a private commercial entity or the press affiliated with another university. When the study appears in a technical journal, the researcher typically has no formal agency relationship with the jour-nal; the researcher simply submits an article documenting the validity study to the journal, and the journal then decides whether to publish the article.

n104 Edward J. Imwinkelried, A Comparativist Critique of the Interface Between Hearsay and Expert Opinion in American Evidence Law, *33 B.C. L. Rev. 1, 15 (1991)*.

n105 *Id. at 15-16*.

n106 Edward J. Imwinkelried, The Use of Learned Scientific Treatises Under Fed-eral Rule of Evidence 803(18), 18 Trial 56 (Feb. 1982).

n107 *Fed. R. Evid. 803(18)*.

n108 Id.

n109 W. Brown Morton, Jr., Recent Development, Medical Treatises to be Admitted as Direct Evidence in Wisconsin—Lewandowski v. Preferred Risk Mutual Ins. Co., *66 Mich. L. Rev. 183, 183 (1967)*.

n110 Notes Learned Treatises, *46 Iowa L. Rev. 463, 466 (1961)*.

n111 *Cal. Evid. Code section 1341*.

n112 1 Bernard Jefferson, California Evidence Benchbook section 18.7, at 435 (2d ed. 1982); Imwinkelried et al., supra note 76 at 310; *People v. Conrad, 107 Cal. Rptr. 421, 430 n. 9 (Cal. Ct. App. 1973)*.

n113 Imwinkelried, supra note 107, at 16.

n114 See supra notes 65-67 and accompanying text.

n115 Joseph L. Peterson et al., Crime Laboratory Proficiency Testing Re-search Program (1978).

n116 Id. at 15, 22, 38.

n117 Id. at 7.

n118 Id. at 28.

n119 Id. at 26.

n120 Peterson et al, supra note 115, at 1.

n121 Id. at 23.

n122 Report of California Association of Crime Laboratory Directors Blind Trial #2, March 29, 1990 (on file in Professor Imwinkelried's office).

n123 Peterson et al., supra note 115.

n124 *Fed. R. Evid. 801(a).*

n125 *Fed. R. Evid. 803(8).*

n126 Id.

n127 5 John H. Wigmore, Evidence in Trials at Common Law section 1633 (Chadbourn rev. 1974).

n128 *Fed. R. Evid. 801(a)*, (c).

n129 Peterson et al., supra note 115, at 23.

n130 Peterson et al., supra note 115, at 23.

n131 *Fed. R. Evid. 805.*

n132 *Fed. R. Evid. 803(6).*

n133 Edward J. Imwinkelried, The Constitutionality of Introducing Evaluative Laboratory Reports Against Criminal Defendants, *30 Hastings L.J. 621, 624-26 (1979)* (collecting cases).

n134 Imwinkelried et al., supra note 103, section 1220 at 339.

n135 Imwinkelried et al., supra note 103, section 1220 at 339.

n136 1 McCormick, Evidence section 208, at 936-37 (4th ed. 1992).

n137 *Fed. R. Evid. 404(b).*

n138 Paul Rothstein, Intellectual Coherence in an Evidence Code, *28 Loy. L.A. L. Rev. 1259, 1265 (1995)*. See also Richard Kuhns, The Propensity to Misunderstand the Character of Specific Acts Evidence, *66 Iowa L. Rev. 777 (1981).*

n139 Black, et al., supra note 11, at 775.

n140 Alexander Pope, Pastoral Poetry and an Essay on Criticism 297 (E. Audra & Aubrey Williams eds., 1961).

n141 *Willis v. Roche Biomedical Laboratories, Inc., 21 F.3d 1368 (5th Cir. 1994),* substituted opinion, *No. Civ. A.H-91-2360 1995 U.S. App. Lexis 20463 (5th Cir. 1995); Stinson v. Physicians Immediate Care Ltd., 63 L.W. 2605* (Ill. App., Mar. 24, 1995).

n142 William L. Prosser & W. Page Keeton, Prosser and Keeton on the Law of Torts section 112, at 788-92 (5th ed. 1984).

n143 See supra note 138 and accompanying text.

n144 Imwinkelried, supra note 80, section 1:03.

n145 *Fed. R. Evid. 404(b).*

n146 *In re Larsen, 59 F.3d 783 (8th Cir. 1995); In re Andrews, 49 F.3d 1404 (9th Cir. 1995); Board of Tr. of Trucking Emp. Pension Fund v. Centra, 983 F.2d 495 (3d Cir. 1992); Herriman v. Conrail, Inc., 887 F. Supp. 1148, 1154 (N.D. Ind. 1995); Brodheim v. Rowland, 783 F. Supp. 1245, 1250 n. 1 (N.D. Cal. 1991),* aff'd in part, vacated in part, *993 F.2d 716 (9th Cir. 1993); In re Daniel M., III, 20 Cal. Rptr. 2d 291, 293 (Cal. Ct. App. 1993);* RRLH, Inc. v. Saddleback Valley Uni-fied *Sch. D., 272 Cal. Rptr. 529 (Cal. Ct. App. 1990); Atlantic Richfield Co. v. State, 262 Cal. Rptr. 683 (Cal. Ct. App. 1989).*

n147 *Mail Order Ass'n of America v. U.S. Postal Service, 986 F.2d 509 (D.C. Cir. 1993).*

n148 Id.; *Ishida v. United States, 59 F.3d 1224, 1230 (Fed. Cir. 1995); In re Oxborrow, 913 F.2d 751 (9th Cir. 1990); Central Montana Elec. v. Adm'r of Bonneville Power, 840 F.2d 1472 (9th Cir. 1988).*

n149 *Brodheim v. Rowland, 783 F. Supp. 1245, 1250 n. 1 (N.D. Cal. 1991),* aff'd in part, vacated in part, *993 F.2d 716 (9th Cir. 1993).*

n150 *Adair v. Troy State University of Montgomery, 892 F. Supp. 1401, 1407 (M.D. Ala. 1995).*

n151 *Fed.R.Evid. 404(b).*

n152 22 Charles A. Wright & Kenneth W. Graham, Jr., Federal Practice and Procedure: Evidence section 5233, at 358 (1978).

n153 Id.

n154 Id.

n155 *Cal. Evid. Code section 175.*

n156 *Fed. R. Evid. 404(b).*

n157 1 McCormick, Evidence section 188 at 792-93 (4th ed. 1992).

n158 *Fed. R. Evid. 405(b).*

n159 Prosser & Keeton, supra note 142, at section 112.

n160 Prosser & Keeton, supra note 142, at section 116.

n161 *Fed. R. Evid. 405.*

n162 Black, et al., supra note 11, at 775.

n163 Victor Gold, Limiting Judicial Discretion to Exclude Prejudicial Evidence, *18 U.C.D. L. Rev. 59, 68-69, 80 (1984).*

n164 *Fed. R. Evid. 404.*

n165 *Fed. R. Evid. 404(a)(1)*-(2).

n166 *Fed. R. Evid. 404(a)(3).*

n167 *Fed. R. Evid. 607* ("The credibility of a witness may be attacked by any party, including the party calling the witness.").

n168 *Fed. R. Evid. 608(a).*

n169 Id.

n170 *Fed. R. Evid. 608(b).*

n171 Id.

n172 *Fed. R. Evid. 609(a).*

n173 *Fed. R. Evid. 609(a)(1).*

n174 *Fed. R. Evid. 609(a)(2).*

n175 *Fed. R. Evid. 608.*

n176 *Fed. R. Evid. 609(a)(2).* See Imwinkelried et al., supra note 103, section 708, at 192-93; Edward Cleary, Preliminary Notes on Reading the Rules of Evidence, *57 Neb. L. Rev. 908, 918-19 (1978).*

n177 Carlson, et al., supra note 12 at 396.

n178 *Committee on Rules of Practice and Procedure of the Judicial Conference of the United States, Preliminary Draft of Proposed Rules of Evidence for the United States District Court and Magistrates, 46 F.R.D. 161, 297 (March 1969).*

n179 Imwinkelried et al., Courtroom Criminal Evidence section 708, at 194 (2d ed. 1993).

n180 Webster's Seventh New Collegiate Dictionary 953, 975 (1972).

n181 Carol Henderson Garcia, Expert Witness Malpractice: A Solution to the Prob-lem of the Negligent Expert Witness, *12 Miss. Coll. L. Rev. 39, 49-52 (1991).*

n182 For example, the Advisory Committee Notes to Rules 607 and 609 contain several references to California authorities. *Fed. R. Evid. 607*, *609* advisory committee's notes.

n183 *Cal. Evid. Code section 786.*

n184 Id. at section 1104.

n185 Parker's Evidence Code of California, supra note 102, at 185.

n186 Parker's Evidence Code of California, supra note 102, at 185.

n187 Ronald J. Allen & Joseph S. Miller, The Common Law Theory of Experts: Deference or Education?, *87 Nw. U. L. Rev. 1131 (1993).*

n188 Id.

n189 See supra notes 1-5 and accompanying text.

n190 *Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993).*

n191 *Id. at 2798* (citing Jack B. Weinstein, *Rule 702 of the Federal Rules of Evi-* dence Is Sound: *It Should Not Be Amended, 138 F.R.D. 631, 632 (1992).*

n192 Edward J. Imwinkelried & Robert G. Scofield, The Recognition of an Accused's Constitutional Right to Introduce Expert Testimony Attacking the Weight of Prosecution Science Evidence: The Antidote for the Supreme Court's Mistaken Assumption in California v. Trombetta, *33 Ariz. L. Rev. 59, 84 (1991)*.

n193 See supra note 49 and accompanying text.

n194 *R. v. Abadom [1983] 1 All E.R. 364* (C.A.).

n195 Jennifer A. James, A Clear Cut Case, *47 Mod. L. Rev. 103, 107 (1984)*.

n196 *Id. at 104*.

n197 Black et al., supra note 11, at 722.

n198 Black et al., supra note 11, at 730.

n199 *Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F.3d 1311, 1315 (9th Cir. 1995),* cert. denied, *116 S. Ct. 189 (1995)*.

n200 *Id. at 1316*.

n201 *Id. at 1315*.

n202 Aldous L. Huxley, Brave New World (1932).

n203 Thomas Huxley, On Medical Education, quoted in The Shorter Bartlett's Familiar Quotations 185 (1959).

n204 See Black et al., supra, note 35.

Galton, Francis. 1892. *Finger Prints* (Buffalo, New York: William S. Hein & Co. Inc., 2003), 100-113.
Reproduced with permission.

Library of Congress Cataloging-in-Publication Data

Galton, Francis, Sir, 1822-1911.
[Finger prints]
Fingerprints / by Francis Galton.
p. cm.
Originally published: Finger prints. London ; New York : Macmillan, 1892.
Includes index.
ISBN 1-57588-742-8 (alk. paper)
1. Fingerprints--Identification. 2. Fingerprints--Classification. I. Title.

GN192 .G24 2003
363.25'8--dc21

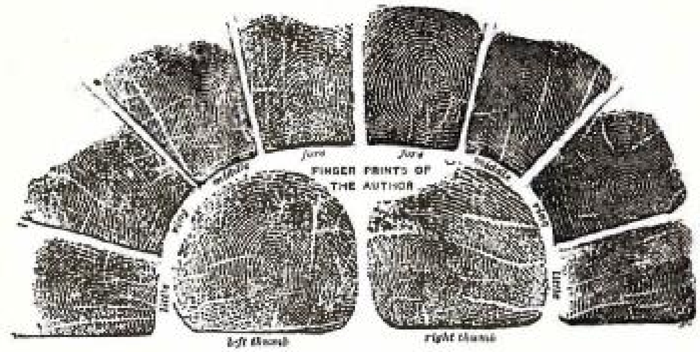
2002038829

This acid-free reprint was electronically imaged and laser printed under the strict quality control guidelines established by William S. Hein & Co., Inc.'s "Preservation Program." This program was established to preserve the integrity of legal classics for future generations of legal researchers.

This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

William S. Hein & Co., Inc.
Buffalo, New York
2003

FINGER PRINTS



BY

FRANCIS GALTON, F.R.S., ETC.

London
MACMILLAN AND CO.
AND NEW YORK
1892

CHAPTER VII

EVIDENTIAL VALUE

The object of this chapter is to give an approximate numerical idea of the value of finger prints as a means of Personal Identification. Though the estimates that will be made are professedly and obviously far below the truth, they are amply sufficient to prove that the evidence afforded by finger prints may be trusted in a most remarkable degree.

Our problem is this: given two finger prints, which are alike in their minutiæ, what is the chance that they were made by different persons?

The first attempt at comparing two finger prints would be directed to a rough general examination of their respective patterns. If they do not agree in being arches, loops, or whorls, there can be no doubt that the prints are those of different fingers, neither can there be doubt when they are distinct forms of the same general class. But to agree thus far goes only a short way towards establishing identity, for the number of patterns that are promptly distinguishable from one another is not large. My earlier inquiries showed this, when endeavouring to sort the prints

of 1000 thumbs into groups that differed each from the rest by an "equally discernible" interval. While the attempt, as already mentioned, was not successful in its main object, it showed that nearly all the collection could be sorted into 100 groups, in each of which the prints had a fairly near resemblance. Moreover, twelve or fifteen of the groups referred to different varieties of the loop; and as two-thirds of all the prints are loops, two-thirds of the 1000 specimens fell into twelve or fifteen groups. The chance that an unseen pattern is some particular variety of loop, is therefore compounded of 2 to 3 against its being a loop at all, and of 1 to 12 or 15, as the case may be, against its being the specified kind of loop. This makes an adverse chance of only 2 to 36, or to 45, say as 2 to 40, or as 1 to 20. This very rude calculation suffices to show that on the average, no great reliance can be placed on a general resemblance in the appearance of two finger prints, as a proof that they were made by the same finger, though the obvious disagreement of two prints is conclusive evidence that they were made by different fingers.

When we proceed to a much more careful comparison, and collate successively the numerous minutiæ, their coincidence throughout would be an evidence of identity, whose value we will now try to appraise.

Let us first consider the question, how far may the minutiæ, or groups of them, be treated as *independent* variables?

Suppose that a tiny square of paper of only one average ridge-interval in the side, be cut out and dropped at random on a finger print; it will mask from view a minute portion of one, or possibly of two ridges. There can be little doubt that what was hidden could be correctly interpolated by simply joining the ends of the ridge or ridges that were interrupted. It is true, the paper might possibly have fallen exactly upon, and hidden, a minute island or enclosure, and that our reconstruction would have failed in consequence, but such an accident is improbable in a high degree, and may be almost ignored.

Repeating the process with a much larger square of paper, say of twelve ridge-intervals in the side, the improbability of correctly reconstructing the masked portion will have immensely increased. The number of ridges that enter the square on any one side will perhaps, as often as not, differ from the number which emerge from the opposite side; and when they are the same, it does not at all follow that they would be continuous each to each, for in so large a space forks and junctions are sure to occur between some, and it is impossible to know which, of the ridges. Consequently, there must exist a certain size of square with more than one and less than twelve ridge-intervals in the side, which will mask so much of the print, that it will be an even chance whether the hidden portion can, on the average, be rightly reconstructed or not. The size of that square must now be considered.

If the reader will refer to Plate 14, in which there are eight much enlarged photographs of portions of different finger prints, he will observe that the length of each of the portions exceeds the breadth in the

proportion of 3 to 2. Consequently, by drawing one line down the middle and two lines across, each portion may be divided into six squares. Moreover, it will be noticed that the side of each of these squares has a length of about six ridge-intervals. I cut out squares of paper of this size, and throwing one of them at random on any one of the eight portions, succeeded almost as frequently as not in drawing lines on its back which comparison afterwards showed to have followed the true course of the ridges. The provisional estimate that a length of six ridge-intervals approximated to but exceeded that of the side of the desired square, proved to be correct by the following more exact observations, and by three different methods.

I. The first set of tests to verify this estimate were made upon photographic enlargements of various thumb prints, to double their natural size. A six-ridge-interval square of paper was damped and laid at random on the print, the core of the pattern, which was too complex in many cases to serve as an average test, being alone avoided. The prints being on ordinary albuminised paper, which is slightly adherent when moistened, the patch stuck temporarily wherever it was placed and pressed down. Next, a sheet of tracing-paper, which we will call No. 1, was laid over all, and the margin of the square patch was traced upon it, together with the course of the surrounding ridges up to that margin. Then I interpolated on the tracing-paper what seemed to be the most likely course of those ridges which were hidden by the square. No. 1 was then removed, and a second sheet, No. 2, was laid on, and the margin of the patch was outlined on it as before, together with the ridges leading up to it. Next, a corner only of No. 2 was raised, the square patch was whisked away from underneath, the corner was replaced, the sheet was flattened down, and the actual courses of the ridges within the already marked outline were traced in. Thus there were two tracings of the margin of the square, of which No. 1 contained the ridges as I had interpolated them, No. 2 as they really were, and it was easy to compare the two. The results are given in the first column of the following table :—

INTERPOLATION OF RIDGES IN A SIX-RIDGE-INTERVAL SQUARE.

Result.	Double Enlargements.	Six-fold scale with prism.	Twenty-fold scale with chequer-work.	Total.
Right .	12	8	7	27
Wrong .	20	12	16	48
Total .	32	20	23	75

II. In the second method the tracing-papers were discarded, and the prism of a camera lucida used. It threw an image three times the size of the photo-enlargement, upon a card, and there it was traced. The same general principle was adopted as in the first method, but the results being on a larger scale, and drawn on stout paper, were more satisfactory and convenient. They are given in the second column of

the table. In this and the foregoing methods two different portions of the same print were sometimes dealt with, for it was a little more convenient and seemed as good a way of obtaining average results as that of always using portions of different finger prints. The total number of fifty-two trials, by one or other of the two methods, were made from about forty different prints. (I am not sure of the exact number.)

The results in each of the two methods were sometimes quite right, sometimes quite wrong, sometimes neither one nor the other. The latter depended on the individual judgment as to which class it belonged, and might be battled over with more or less show of reason by advocates on opposite sides. Equally dividing these intermediate cases between "right" and "wrong," the results were obtained as shown. In one, and only one, of the cases, the most reasonable interpretation had not been given, and the result had been wrong when it ought to have been right. The purely personal error was therefore disregarded, and the result entered as "right."

III. A third attempt was made by a different method, upon the lineations of a finger print drawn on about a twenty-fold scale. It had first been enlarged four times by photography, and from this enlargement the axes of the ridges had been drawn with a five-fold enlarging pantagraph. The aim now was to reconstruct the entire finger print by two successive and independent acts of interpolation. A sheet of transparent tracing-paper was ruled into six-ridge-interval squares, and every one of its alternate squares was rendered opaque by pasting white paper upon it, giving it the appearance of a chess-board. When this chequer-work was laid on the print, exactly one half of the six-ridge squares were masked by the opaque squares, while the ridges running up to them could be seen. They were not quite so visible as if each opaque square had been wholly detached from its neighbours, instead of touching them at the extreme corners, still the loss of information thereby occasioned was small, and not worth laying stress upon. It is easily understood that when the chequer-work was moved parallel to itself, through the space of one square, whether upwards or downwards, or to the right or left, the parts that were previously masked became visible, and those that were visible became masked. The object was to interpolate the ridges in every opaque square under one of these conditions, then to do the same for the remaining squares under the other condition, and finally, by combining the results, to obtain a complete scheme of the ridges wholly by interpolation. This was easily done by using two sheets of tracing-paper, laid in succession over the chequer-work, whose position on the print had been changed meanwhile, and afterwards tracing the lineations that were drawn on one of the two sheets upon the vacant squares of the other. The results are given in the third column of the table.

The three methods give roughly similar results, and we may therefore accept the ratios of their totals,

which is 27 to 75, or say 1 to 3, as representing the chance that the reconstruction of any six-ridge-interval square would be correct under the given conditions. On reckoning the chance as 1 to 2, which will be done at first, it is obvious that the error, whatever it may be, is on the safe side. A closer equality in the chance that the ridges in a square might run in the observed way or in some other way, would result from taking a square of five ridge-intervals in the side. I believe this to be very closely the right size. A four-ridge-interval square is certainly too small.

When the reconstructed squares were wrong, they had none the less a natural appearance. This was especially seen, and on a large scale, in the result of the method by chequer-work, in which the lineations of an entire print were constructed by guess. Being so familiar with the run of these ridges in finger prints, I can speak with confidence on this. My assumption is, that any one of these reconstructions represents lineations that might have occurred in Nature, in association with the conditions outside the square, just as well as the lineations of the actual finger print. The courses of the ridges in each square are subject to uncertainties, due to petty *local* incidents, to which the conditions outside the square give no sure indication. They appear to be in great part determined by the particular disposition of each one or more of the half hundred or so sweat-glands which the square contains. The ridges rarely run in evenly flowing lines, but may be compared to footways across a broken country, which, while they follow a general direction, are continually deflected by such trifles as a tuft of grass, a stone, or a puddle. Even if the number of ridges emerging from a six-ridge-interval square equals the number of those which enter, it does not follow that they run across in parallel lines, for there is plenty of room for any one of the ridges to end, and another to bifurcate. It is impossible, therefore, to know beforehand in which, if in any of the ridges, these peculiarities will be found. When the number of entering and issuing ridges is unequal, the difficulty is increased. There may, moreover, be islands or enclosures in any particular part of the square. It therefore seems right to look upon the squares as independent variables, in the sense that when the surrounding conditions are alone taken into account, the ridges within their limits may either run in the observed way or in a different way, the chance of these two contrasted events being taken (for safety's sake) as approximately equal.

In comparing finger prints which are alike in their general pattern, it may well happen that the proportions of the patterns differ; one may be that of a slender boy, the other that of a man whose fingers have been broadened or deformed by ill-usage. It is therefore requisite to imagine that only one of the prints is divided into exact squares, and to suppose that a reticulation has been drawn over the other, in which each mesh included the corresponding parts of the former print. Frequent trials have shown that there is no practical difficulty in actually doing

this, and it is the only way of making a fair comparison between the two.

These six-ridge-interval squares may thus be regarded as independent units, each of which is equally liable to fall into one or other of two alternative classes, when the surrounding conditions are alone known. The inevitable consequence from this datum is that the chance of an exact correspondence between two different finger prints, in each of the six-ridge-interval squares into which they may be divided, and which are about 24 in number, is at least as 1 to 2 multiplied into itself 24 times (usually written 2^{24}), that is as 1 to about ten thousand millions. But we must not forget that the six-ridge square was taken in order to ensure under-estimation, a five-ridge square would have been preferable, so the adverse chances would in reality be enormously greater still.

It is hateful to blunder in calculations of adverse chances, by overlooking correlations between variables, and to falsely assume them independent, with the result that inflated estimates are made which require to be proportionately reduced. Here, however, there seems to be little room for such an error.

We must next combine the above enormously unfavourable chance, which we will call a, with the other chances of not guessing correctly beforehand the surrounding conditions under which a was calculated. These latter are divisible into b and c; the chance b is that of not guessing correctly the general course of the ridges adjacent to each square, and c that of not guessing rightly the number of ridges that enter and issue from the square. The chance b has already been discussed, with the result that it might be taken as 1 to 20 for two-thirds of all the patterns. It would be higher for the remainder, and very high indeed for some few of them, but as it is advisable always to underestimate, it may be taken as 1 to 20; or, to obtain the convenience of dealing only with values of 2 multiplied into itself, the still lower ratio of 1 to 2^4, that is as 1 to 16. As to the remaining chance c with which a and b have to be compounded, namely, that of guessing aright the number of ridges that enter and leave each side of a particular square, I can offer no careful observations. The number of the ridges would for the most part vary between five and seven, and those in the different squares are certainly not quite independent of one another. We have already arrived at such large figures that it is surplusage to heap up more of them, therefore, let us say, as a mere nominal sum much below the real figure, that the chance against guessing each and every one of these data correctly is as 1 to 250, or say 1 to 2^8 ($=256$).

The result is, that the chance of lineations, constructed by the imagination according to strictly natural forms, which shall be found to resemble those of a single finger print in all their minutiæ, is less than 1 to $2^{24} \times 2^4 \times 2^8$, or 1 to 2^{36}, or 1 to about sixty-four thousand millions. The inference is, that as the number of the human race is reckoned at about sixteen thousand millions, it is a smaller chance than 1 to 4 that the print of a *single* finger of any given person would be exactly like that of the same finger of any other member of the human race.

When two fingers of each of the two persons are compared, and found to have the same minutiæ, the improbability of 1 to 2^{36} becomes squared, and reaches a figure altogether beyond the range of the imagination; when three fingers, it is cubed, and so on.

A single instance has shown that the minutiæ are *not* invariably permanent throughout life, but that one or more of them may possibly change. They may also be destroyed by wounds, and more or less disintegrated by hard work, disease, or age. Ambiguities will thus arise in their interpretation, one person asserting a resemblance in respect to a particular feature, while another asserts dissimilarity. It is therefore of interest to know how far a conceded resemblance in the great majority of the minutiæ combined with some doubt as to the remainder, will tell in favour of identity. It will now be convenient to change our datum from a six-ridge to a five-ridge square of which about thirty-five are contained in a single print, 35×5^2 or 35×25 being much the same as 24×6^2 or 24×36. The reason for the change is that this number of thirty-five happens to be the same as that of the minutiæ. We shall therefore not be acting unfairly if, with reservation, and for the sake of obtaining some result, however rough, we consider the thirty-five minutiæ themselves as so many independent variables, and accept the chance now as 1 to 2^{35}.

This has to be multiplied, as before, into the factor of $2^4 \times 2^8$ (which may still be considered appropriate, though it is too small), making the total of adverse chances 1 to 2^{47}. Upon such a basis, the calculation is simple. There would on the average be 47 instances, out of the total 2^{47} combinations, of similarity in all but one particular; $\frac{47\times46}{1\times2}$ in all but two; $\frac{47\times46\times45}{1\times2\times3}$ in all but three, and so on according to the well-known binomial expansion. Taking for convenience the powers of 2 to which these values approximate, or rather with the view of not overestimating, let us take the power of 2 that falls short of each of them; these may be reckoned as respectively equal to 2^6, 2^{10}, 2^{14}, 2^{18}, etc. Hence the roughly approximate chances of resemblance in all particulars are as 2^{47} to 1; in all particulars but one, as 2^{47-6}, or 2^{41} to 1; in all but two, as 2^{37} to 1; in all but three, as 2^{33} to 1; in all but four, as 2^{29} to 1. Even 2^{29} is so large as to require a row of nine figures to express it. Hence a few instances of dissimilarity in the two prints of a single finger, still leave untouched an enormously large residue of evidence in favour of identity, and when two, three, or more fingers in the two persons agree to that extent, the strength of the evidence rises by squares, cubes, etc., far above the level of that amount of probability which begins to rank as certainty.

Whatever reductions a legitimate criticism may make in the numerical results arrived at in this chapter, bearing in mind the occasional ambiguities pictured in Fig. 18, the broad fact remains, that a complete or nearly complete accordance between two prints of a single finger, and vastly more so between

the prints of two or more fingers, affords evidence requiring no corroboration, that the persons from whom they were made are the same. Let it also be remembered, that this evidence is applicable not only to adults, but can establish the identity of the same person at any stage of his life between babyhood and old age, and for some time after his death.

We read of the dead body of Jezebel being devoured by the dogs of Jezreel, so that no man might say, "This is Jezebel," and that the dogs left only her skull, the palms of her hands, and the soles of her feet; but the palms of the hands and the soles of the feet are the very remains by which a corpse might be most surely identified, if impressions of them, made during life, were available.

Wertheim, Kasey, and Alice Maceo, “The Critical Stage of Friction Ridge Pattern Formation” (2002) 52(1) Journal of Forensic Identification 35-84. Courtesy of the International Association for Identification. Reproduced with permission.

Article

The Critical Stage of Friction Ridge and Pattern Formation

***Kasey Wertheim*[1]**
***Alice Maceo*[2]**

Abstract

This study provides an enhanced understanding of the biological structure and development of friction ridge skin for the latent print examiner who is called upon to explain the scientific principles of latent print identification as based on permanence and uniqueness. Cellular attachments ensure permanence, while variable stresses and cellular distributions account for individuality on all "three levels" of detail. Volar patterning is dependent upon the tension across the surface of the developing skin during a critical stage of approximately 10.5 to 16 weeks estimated gestational age. Fingerprint ridge counts are predominantly affected by two combined timing events: the onset of epidermal cellular proliferation and the timing of the regression of the volar pads. Fingerprint pattern types are predominantly affected by the symmetry of the volar pad.

Introduction

The accuracy and reliability of many of the forensic sciences are currently being challenged. Specific among these challenges is the admissibility and reliability of the science related to friction ridge skin identification. All latent print examiners should understand and be prepared to recite the philosophy of identification, define the methodology used in examinations, and explain the scientific basis of the permanence and uniqueness of friction ridge skin based upon empirical and scientific research data.

---

1 - Latent Print Examiner, Mississippi Crime Laboratory. Meridian, MS
2 - Latent Print Examiner, Henderson Police Department, Henderson, NV

In the United States, legal challenges relating to the admissibility of friction ridge skin identifications can occur on a case-by-case basis. Until 1993, the standard of admissibility was that the scientific technique or procedure be generally accepted and practiced by the relevant scientific community. *Daubert vs. Merrell Dow Pharmaceuticals* (1993) enhanced the authority of trial judges by giving them "gatekeeper" authority to alleviate possible "junk science" entering the courtroom [1]. Although Daubert issues have surfaced during the trial phase. traditionally, a "Daubert hearing" is conducted pre-trial and outside the presence of the jury. In addition to the relevancy of the subsequent case testimony, several factors may be considered by the trial judge in making a reliability determination regarding scientific principles or methodology including: (1) testing and validation, (2) publication with peer review. (3) error rate, (4) existence of standards, and *(5)* general acceptance.

Forensic sciences have inadvertently stepped into the limelight of prime time newscasts and television dramas, sparking public interest and debate. Latent print examiners are being questioned about their science not only in legal settings, but in social and public venues as well. As the international community grapples with updating policies and procedures to meet new court challenges, it must also address the concerns of the general public.

It is of the utmost importance that each examiner in the field of latent print examination be able to explain why friction ridge skin is unique and permanent. The traditional response, "because no two fingerprints have ever been found to be the same," is inadequate. Knowledge of the physical and natural scientific basis, rather than observational data alone, must be exhibited to explain why friction ridge skin is unique and why this uniqueness is permanent. This study explores the anatomical structure and growth process of human friction ridge (volar) skin presented in a manner to enhance latent print examiner understanding, and equip the latent print examiner with scientific information to accurately and concisely communicate the principles of the science of friction ridge skin identification.

The Structure of Friction Skin

Like all skin, volar skin is composed of two basic layers: the outer epidermis, and the inner dermis (Figure 1). The epidermis and dermis are separated by a basement membrane, which serves as a boundary and a mechanical linkage between these two tissue layers. The epidermis is further divided into five cellular layers based on intrinsic changes in the cells as they progress from the bottom of the epidermis to the surface of the skin. The dermis is divided into two layers, reflecting differences in fiber composition, cell type and distribution, and vascular networks [2].

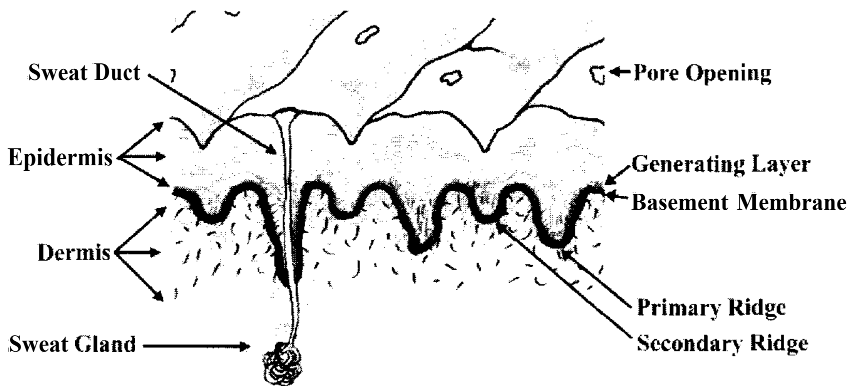


Figure 1

Three dimensional representation of the structure of mature volar skin.

The surface ridges and furrows of friction skin reflect, to a certain degree, the complex organization of the epidermis below the surface. The basal layer of the epidermis of friction skin has a series of folds protruding into the dermis, which correspond to the ridges and furrows on the outer surface of the epidermis. The ridges or folds of the basal layer containing ducts from the eccrine sweat glands of volar skin are termed primary ridges, and correspond to the surface ridges of friction skin. Secondary ridges, alternating between primary ridges, also protrude into the dermis, but correspond to the furrows on the surface of the skin.

Branches of epidermis, termed “anastomoses” bridge primary and secondary ridges and mold the papillae pegs of the dermis [3]. The series of primary and secondary ridges, accentuated by the anastomoses, increase the surface area of attachment between the epidermis and dermis. New epidermal cells are constantly generated in the basal layer and pushed toward the surface [2]. These new cells replenish the upper layers of epidermis and consistently reproduce the complex arrangements present at the epidermal-dermal junction.

The structure of volar skin is extremely complex, and is very difficult to represent in a two-dimensional drawing. Available literature contains representations of the structure of friction skin that, on first glance. appear to conflict with regard to primary, secondary, and surface ridge configuration. In some drawings of the structure of fetal skin, "double rows'' of dermal papillae appear to correspond with surface furrows (Figure 2A), [4, 5, 6] whereas other drawings show the double rows as residing beneath the surface ridge itself (Figure 2B) [7, 8]. However. these apparent diagramatical discrepancies can be resolved when the growth of friction ridge skin is viewed with the added consideration of the passage of time (Figures 2C and 2D). Further, Chacko noted skin of different structure in different individuals of the same age, as well as in different areas of friction ridge skin of the same person, designating them as types "DR0, DRI, DR2. and DR3" (Figure 2) [9].

The Principle of Permanence

The key to understanding why friction skin retains its features throughout natural growth and aging lies within the structure of skin. There are three principal structural elements of skin that allow for the permanence of friction ridge detail: (1) the adherence of the epidermal cells to each other, (2) the basal cell layer of the epidermis, and its attachment to the basement membrane. and (3) the attachment of the basement membrane to the dermis. A general discussion on the cellular structure of skin is included to demonstrate the specific attributes that enforce permanence.

Epidermis

Epidermis of volar skin is approximately 1.8mm in thickness [8], and is comprised mostly of cells called *keratinocytes.* The keratinocyte undergoes a maturation process, termed differentiation [2], as it moves from the basal layer of the epidermis to the surface, where it sloughs off (exfoliates) into the atmosphere. The term *differentiation* represents the progression of the keratinocyte from a newly generated cell, through synthesis and accumulation of the protein keratin, and, finally, to a dead, completely keratinized (cornified) cell [10].

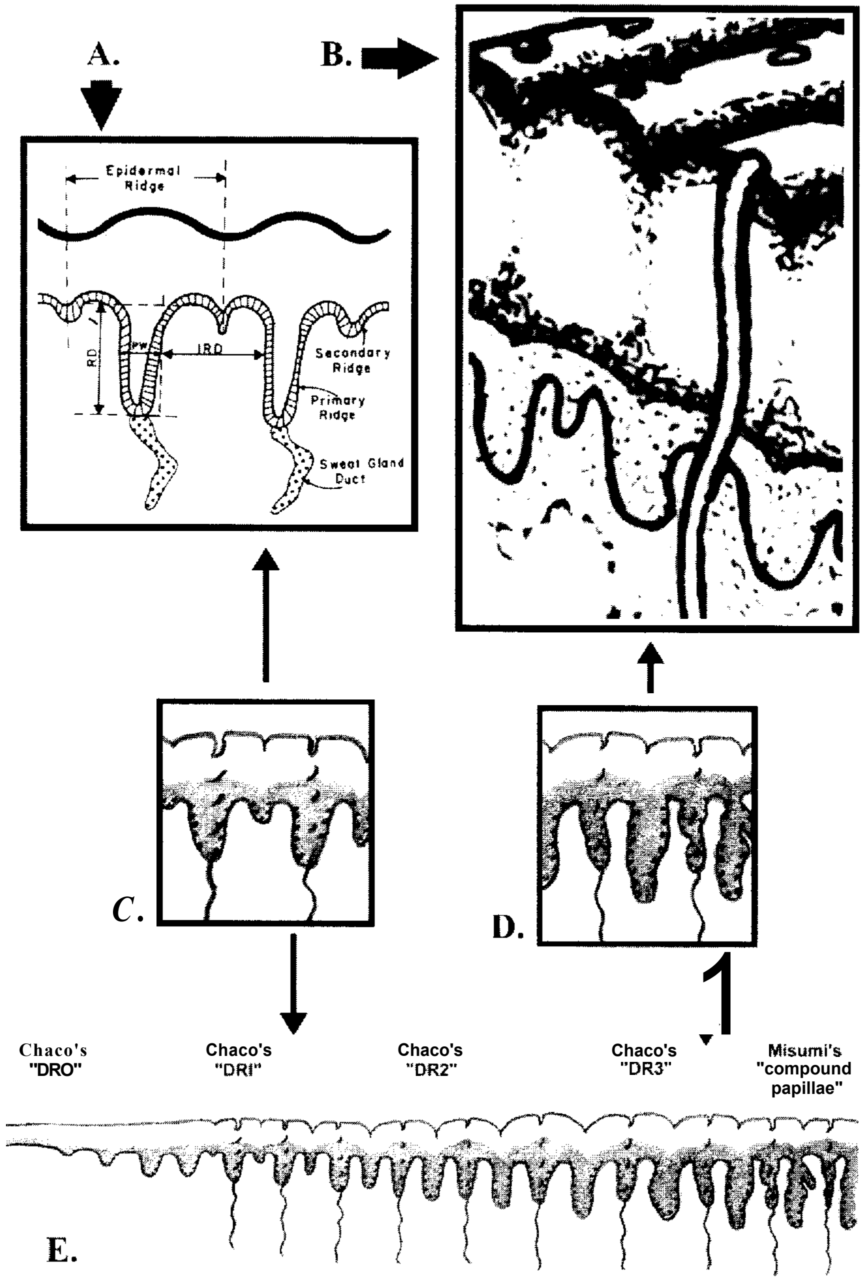


Figure 2

A: Drawing of fetal volar skin structure [4] B: Drawing of adult volar **skin** *structure [7] C: A section of E which corresponds with A D: A section of E which corresponds with B E: An illustration of the progression of volar skin from DRO (left) through DR3 fright) [9] ana' representing papillae of different sizes and shapes (right) [5]*

The *stratum germinativum* (generating layer) is the basal layer of keratinocytes of the epidermis. It is a single layer of columnar shaped cells attached to the basement membrane. These cells undergo mitosis (cell division) and, because of the direction of division in these columnar cells, the new cells are pushed upward within the epidermis to replace exfoliated cells. Basal cells are joined to each other by complex cell junctions called *desmosomes* and are joined to the basement membrane by another type of junction called a *hemidesmosome* [2].

After dividing from the basal cells of the generating layer, keratinocytes are pushed upward and become part of the *stratum spinosum* (spinous layer), a zone 2 to 4 cell layers in thickness. Cells of the spinous layer exhibit the first stage of differentiation by organizing the necessary components of keratin synthesis. These cells are bound to one another by abundant desmosomes [2].

As cells progress toward the external surface of the epidermis, they become part of the *stratum granulosum* (granular layer). The cells of this layer are the last of the living cells of the epidermis, as they begin to display the first precursors of keratin. As keratinization occurs, all other cellular activities and components degrade, marking the beginning stages of cell death. Some researchers define the small layer of histologically clear cells at the top of the granular layer as the *stratum lucidum*, or Hyalin layer [11]. The cells remain linked by desmosomes as they progress into the *stratum corneum* (horny layer) [2].

By the time the cells have reached the horny layer of the upper epidermis, they have accumulated keratin, and cell death has occurred. The horny layer of adult volar skin can be up to 100 cells in thickness [2], making it 25 to 30 times thicker than most other areas of the body [12]. Cornified, dead cells are large and flat (Figure 3), overlap at the margins, and are joined to underlying and superficial layers through interlocking undulations and modified desmosomes [2]. The entire progression through the different stages of cell maturation, from cell birth to exfoliation, is approximately 30 days [12]. Individuals who live to be 60 years old will undergo complete turnovers of epidermis approximately 720 times in their lifetime. This fact reinforces the need for latent print examiners to understand why ridge detail is consistently represented on the surface of the skin, year after year.

The first structural element that enforces the permanence of friction ridges (and other skin features) is the complex and secure junction between cells of the epidermis. Cells generated from the basal layer are surrounded by and "cemented" to neighboring cells in the epidermis, and remain so until exfoliation occurs [11, 13]. The positional properties of the basement membrane zone are consistently transferred to the surface because individual cells move upward in concert with surrounding cells.

Figure 3

Scanning electron view of the surface of a friction ridge showing cornified epidermal cells undergoing exfoliation, and the orifice of a sweat gland. [11]

Basement Membrane

The second structural element of permanence is the attachment site between the basal cells of the epidermis and the basement membrane. Cell junctions and small fibers prevent the basal cells from sliding or migrating along the basement membrane. Further, the rate at which a basal cell produces new epidermal cells is reasonably consistent. For normal healthy tissue, a small isolated area of basal cells will not suddenly begin to produce more or less epidermal cells than it previously did, thereby causing a different formation on the surface ridge. This concept is supported by 100 years of observational data demonstrating that ridge detail is not subject to variations over time. barring growth, disease or injury.

Dermis

The dermis lies directly beneath the basement membrane and is anchored to it by penetrating fibrils and bundles of microfibrils. The dermis is divided into two regions, the upper papillary dermis and the lower reticular dermis. Both regions of the dermis are composed of fibrous and nonfibrous dermal matrix. The fibrous material (primarily collagen and elastic fibers) gives the skin bulk and tensile strength while allowing for flexibility. The nonfibrous material forms a ground substance which influences the passage of nutrients, allows for cellular migration, and provides a continuous medium for structural fibers [2].

The third structural element attributing to permanence is the attachment of the dermis to the basement membrane through small, anchoring fibers. The extensive network of fibrils and microfibrils prevents the dermis from sliding along the basement membrane.

Basement Membrane Zone Properties, Aging, and Third Level Detail

The epidermal-dermal junction is referred to as the basement membrane zone (BMZ). The BMZ is composed of constituents from the epidermis and the dermis and separates the two layers. In addition to its structural function, the BMZ acts as a filter between the epidermis and dermis. **All** nutrients, waste, and chemical signals to and from the epidermis must pass through the basement membrane zone [2].

The epidermal-dermal junction becomes more complex with age. Dermal papillae increase in number and become arranged in a progressively more crowded state throughout adulthood [5, 9, 14]. These changes most likely reflect the continued addition of anastomoses between primary and secondary epidermal ridges, providing enhanced attachment sites for aging skin.

It is important for the latent print examiner to realize why changes in papilla configurations do not affect the detail represented on the surface of the skin. Certain concepts regarding the basement membrane zone must be presented and understood on a cellular level to address this issue.

It is easy to visualize a three-dimensional structure beneath the surface that exactly mirrors surface ridges, but this is not entirely accurate. Surface ridge configurations, down to third level detail (ridge and pore shape), are not solely related to a similar shape along the basement membrane. They are also rooted in the configuration and type of basal cells which feed new epidermal cells to the surface. These are very different concepts.

In an effort to differentiate the concepts, imagine a row of 10 devices and that each rolls a marble across a table at a constant interval. Further imagine in this analogy that each of these devices rolls marbles at different rates: one marble per second, two per second, three per second, etc. At the end of the table are 10 baskets that collect the marbles from each device. If the devices are started at the same time and stopped 5 minutes later, the number of marbles in each basket will be different. If the devices were misaligned to slightly different distances from the edge of the table, and the experiment were re-run, the number of marbles in each basket after 5 minutes would not be significantly different than the previous run; perhaps only by a marble or two. On the other hand, if the production rates of the devices were changed, large differences would be expected in the numbers of marbles in each respective basket. The rates of production of the devices have a much more drastic effect on the end result than simply changing their position on the table.

When visualizing the three-dimensional epidermal-dermal junction, whether an area is concave or convex does not appear to be as important as the basal cell types present and the mitosis rates of the different (heterogeneous) basal cells in determining the three-dimensional detail on the surface within that particular area. The devices in our analogy mimic basal cells, which do not supply new cells at the same rate [2, 13]. Stem cells directly underneath primary ridges (and not secondary ridges) in monkey palms have been shown to give rise to "transient amplifying cells" which, themselves, "undergo a few rounds of cell division" before differentiation [15]. It is important to understand that even though differences in the rate of cell proliferation may occur over time [13], those changes affect relatively large areas of epidermis, therefore, a particular detail in an isolated area would not change. In short, basal cells constantly produce new cell growth at rates proportional to surrounding basal cells [3].

Although latent print examiners may not be able to relate the scientifically-technical details of the mechanisms of human basal cell mitosis, they should be able to demonstrate in lay analogies the complex nature of the epidermal-dermal junction. This communication would explain why third level detail remains permanent during the aging process of skin, in spite of changes in the shape of the epidermal-dermal boundary.

The Principle of Individuality

The individuality of friction ridge skin, or all skin for that matter [13], falls under the larger umbrella of biological uniqueness. No two organisms are exactly alike. The intrinsic and extrinsic factors affecting the development of any individual are impossible to duplicate. The most obvious example is the fact that monozygotic twins are each unique, distinguishable individuals.

Just as homes built from the same blueprint are not the same house. individuals with identical DNA are not the same person. The same blueprint can be used to build six homes, but the final outcome of each home depends on available materials, manner of construction, and a host of environmental factors too numerous to even fathom. Although the six homes will be similar, they will not be exactly alike. The boards of the wood frame will not be in the same exact position. the thousands of nails and screws will not be in the same exact locations, and the rooms will not have exactly the same dimensions.

DNA works in a very similar way. It provides a blueprint for assembling proteins. These proteins direct the cell's activities by facilitating biochemical processes within the cell. These processes not only depend on the protein derived from the gene, but also the many other components of the cell such as sugars, lipids, non-protein hormones, inorganic elements (e.g. oxygen), inorganic compounds (e.g. nitric oxide), and minerals. Additionally, the physical environment around and within cells, such as surface tension, electrical charge, and viscosity, contribute to the way the cell functions [16].

Genetic information directs cellular function, serves as a link between generations, and influences an individual's appearance. Some aspects of appearance are similar for each individual of that species (i.e. those characteristics which define the species). However, within the species, for each aspect of an individual's appearance. there are many genes and external factors that affect the final outcome of physical appearance. The genes involved with a specific attribute (e.g. skin color) produce the appropriate proteins. which in turn react with the many non-genetic components of the cell and react with each other in complex biochemical pathways during the growth and development of the fetus [16]. These biochemical pathways proceed under the omnipresent influence of external factors.

Although DNA is crucial for providing the blueprint for the development of a particular model, there are so many steps between the genesis of the DNA-encoded protein and the final product, that even the same DNA blueprint produces two completely unique models. With respect to fingerprint patterns, one milestone along the pathway from blueprint to model involves the development of the central nervous and cardiovascular systems.

Trigger Mechanism for the Onset of Friction Ridge Proliferation

At the time of embryonic friction ridge formation, the central nervous and cardiovascular systems are undergoing a critical period of development [17]. Many researchers have reported the appearance of nerve endings (innervation) at the sites of ridge formation immediately preceding the appearance of ridges, and suggest this could be the trigger mechanism for the onset of cell division (proliferation) [18, 19, 20. 21]. Several researchers even postulated that the distribution of the capillary-nerve pairs at the junction of the epidermis and dermis *directly* influences the alignment of the primary ridges [19, 20, 21]. Earlier research on pattern distribution established "developmental fields," or groupings of fingers on which patterns had a greater tendency to be similar [22, 23, 24]. Later discoveries confirm the neurological relation of spinal cord sections C-6, C-7, and C-8 to innervation of the fingers [25]. This offers even more support of the link betw-een innervation and volar patterning (dermatoglyphics).

The presence of nerves and capillaries in the dermis prior to friction ridge formation may be necessary for friction ridge proliferation. It would seem that all relevant areas of the developing fetus must be in communication with the central nervous system or the endocrine and exocrine (hormone) systems in order to orchestrate complex simultaneous productions such as friction ridge formation [26]. However, it is doubtful that the nerves and/or capillaries independently establish a map that directly determines subsequent ridge flow. It seems more likely that the alignment of the nerves and/or capillaries is directed by the same stresses and strains of the developing hand that establish ridge alignment [26, 27].

Another theory proposes that the same forces of compression on the deeper layers of the epidermis which condition ridge alignment also stimulate the proliferation of basal cells [28]. It is well recognized in cell biology that physical pressure on a cellular system can trigger electro-chemical changes within that system itself. Merkell cells occupy the epidermis just prior to innervation along those pathways [2], suggesting that even this early in fetal formation the stresses created by the different growth rates of the dermis and epidermis are causing cellular activity along invisible lines of stress which already delineates pattern characteristics. Regardless of the trigger mechanism controlling the onset of the first primary ridge proliferations, the structure of the skin during this critical stage of development has been well documented.

Primary Ridge Structure and Genesis

Prior to ridge development, the embryonic epidermis is three to four cell layers thick, and smooth on its outer surface (periderm) and its inner surface (dermal-epidermal junction). Keratinocytes are tightly bound to each other by desmosomes and the cells of the basal layer are attached to the basement membrane by hemidesmosomes [2]. At around 10 to 10.5 weeks Estimated Gestational Age (EGA), basal cells of the epidermis begin to rapidly divide [29]. As the cells proliferate, shallow "ledges" [3] can be seen on the bottom of the epidermis which already delineate the overall patterns that will become permanently established on the volar surfaces several weeks later [29, 30].

During embryonic development, primary ridges are the first visual evidence of interaction between the dermis and epidermis, and arc first seen forming as continuous ridges (Figure 4). The prevailing theory of events prior to the visualization of primary ridge structure involves centers of active cell proliferation (Figure 5), which will be the center of sweat gland development [3]. Under this theory. the "units" of rapidly multiplying cells increase in diameter, somewhat randomly growing into one another (Figure **6)** along the lines of stress and strain relief that run perpendicularly to the direction of tension. As the series of localized proliferations "fuse" together, the resulting linear ridges of rapidly dividing epidermal cells fold into the dermis, creating the first visible ridge structure at the epidermal-dermal junction [13].

In 1904, Inez Whipple presented research detailing a theory of evolutionary progression of the volar surface. Ashbaugh succinctly summarizes Whipple's proposition of the evolutionary genesis of friction ridges:

> "Early mammals were covered with a scale-like skin surface. Each scale had one hair protruding from it and an accompanying oil or sebaceous gland. On volar areas, which are the bottoms of the hands and feet, hairs slowly disappeared due to surface use. The pore that was related to the hair changed from a sebaceous gland to a sweat gland. Its purpose, to keep the surface skin damp which enhanced the grip of the volar surface.
>
> Starting in all likelihood as a mutation, scales started to line up in rows and fuse together. This further assisted the

grip of the skin surface by increasing friction. Through natural selection, this mutation became prevalent. Scales slowly evolved into wart-like units with pore openings near the centre. The fusing of these wart formations into rows is the predcccssor to thc friction ridge, the individual wart being the equivalent of a ridge dot." [31]

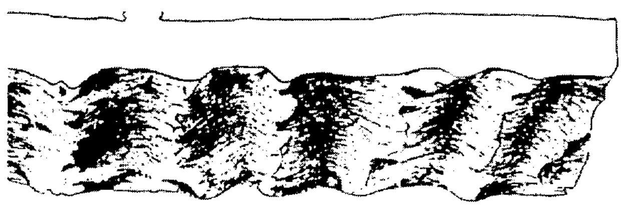


Figure 4

Reconstruction ofthe underside of fetal volar epidermis, displaying the first three-dimensional model of the undulations ut the epidermal-dermal junction [3].

Figure 5

Histologic. cross section of 10.5 week EGA fetal volar. skin at the onset of cellular proliferation [29].

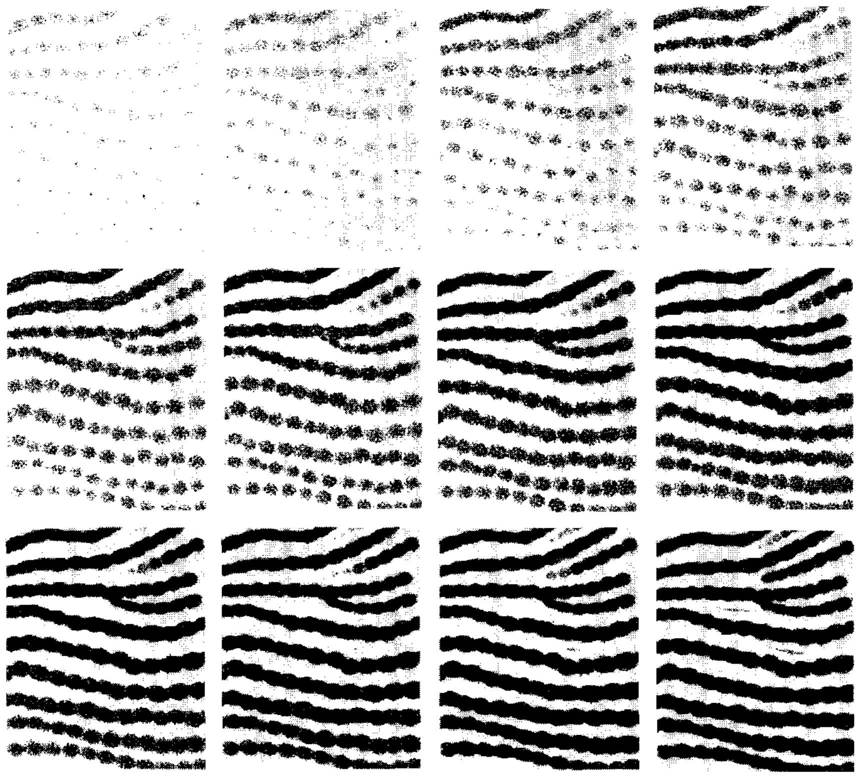


Figure 6

This series of drawings represent the currently accepted theory that localized cellular proliferations grow together into what will subsequently appear as ridges at the epidermal-dermal junction ut around 10.5 weeks EGA.

Thirteen years after Whipple's *phylogenetic* (evolutionary history) theory was presented, researchers diverged from her theory and presented an *ontogenetic* (developmental or embryonic history of an individual) model, suggesting that fusion of warts into ridges occurs during the embryonic development process [32]. In 1926, Cummins refuted the ontogenetic scheme [33]. However, Hale later included the ontogenetic model in his conclusions [3]. Literature since that time has been mixed. The fact remains that, currently, the first visual evidence of interaction between the dermis and the epidermis clearly demonstrates ridges forming as ridges, not a series of units protruding into the dermis (Figure 4). Perhaps with advances in technology, the theory that localized cell proliferations grow together into linear ridges before the appearance of the structure as a ridge will be demonstrated. Until then, this will have to remain a possible model of development which could provide individuality prior to the appearance of the first ridge structures. The term "Ridge Unit" might be limited to a description of an adult sweat pore and surrounding ridge [13], with the term "Localized Proliferation" being used to describe theoretical events of fetal formation [34].

Primary ridges mature and extend deeper into the dermis. Initial formation of primary ridges begins at about 10.5 weeks EGA and continues until about 16 weeks EGA, at which time secondary ridges begin to form. Although the exact mechanisms by which minutia (Galton's details, second level details) form is unclear, observational data by many researchers examining fetal tissue provides a detailed visual account of the structure of friction ridge skin in successive stages of the development process. The general consensus of the literature in detailing the formation of second level detail is represented in Figure 7. Many things are happening during this period of primary ridge growth. The finger is growing, new primary ridges are forming across the finger, and the existing primary ridges are beginning to separate due to growth of the digit. As existing ridges separate. a demand for new ridges is created because the surface has a tendency to be continually ridged at this point in development. New ridges pull away from existing primary ridges [3] to fill in these gaps, creating bifurcations by mechanical separation. Ending ridges form when a developing ridge becomes sandwiched between two established ridges. Under this theory, "fusion between adjacent ridges [which have already formed] seems improbable, although there is no evidence for or against this process." [3]

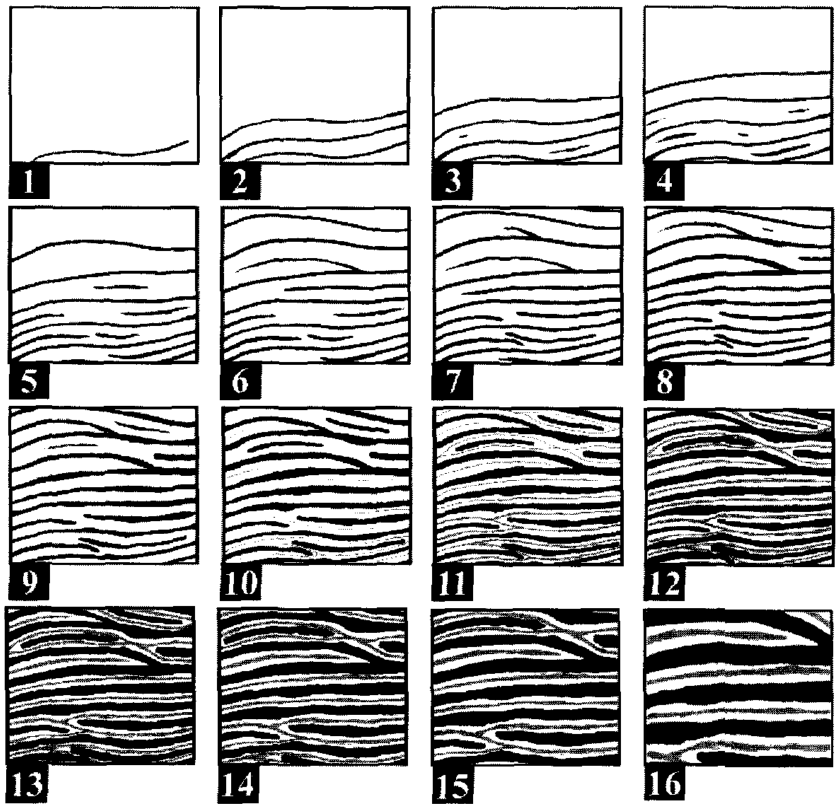


Figure 7

This series of drawings represents the consensus of the literature in demonstrating the theoreticaljormation of niinutia arising from expansion ofthe volar surface and the tendency of volar skin during the critical stage (frames I - 10) to remain continuously ridged. Once secondary ridge formation begins at about 16 weeks EGA (frame 10), the minutia becomes set and the ridges will onlv increase in size during maturity.

An alternate theory is that all minutia form within the pattern from the onset of proliferation, remain transient during the critical stage. and become permanently set upon secondary ridge formation. The difference between the two theories is subtle: bifurcations would occur as a mechanical separation in the first theory versus a static formation in the second. Under this static theory, both bifurcations and ridge endings could form as a result of the random fusion of the localized cell proliferations, or by another mechanism such as chemical reaction-suppression models.

Regardless of which theory of minutia formation is considered (mechanical or static, fusion or chemical). the placement of any particular second level detail within the developing ridge field is governed by a random series of infinitely interdependent stresses, strains, and tensions across that particular area of skin at that critical moment. Slight differences in the mechanical stress, physiological environment andior variation in the timing of development could affect any particular minutia placement in that area of skin.

Secondary Ridge Formation

Sweat glands begin to appear around 14 weeks EGA as the existing primary ridges increase in width and continue to penetrate the dermis [29]. By 15 weeks EGA, the primary ridges are experiencing development in two directions: the downward penetration of the sweat glands and the upward push of new cell growth. Between 15 and 17 weeks EGA, secondary ridges appear between the primary ridges (Figure 8) on the underside of the epidermis [29]. At this point in fetal development, the randomly located minutia within the fingerprint pattern become permanently set [3], marking the end of new primary ridge formation [35] and the end of the critical stage.

Secondary ridges are also cell proliferations resulting in downfolds of the basal epidermis. As the secondary ridges form downward and increase the surface area of attachment to the dermis, the primary ridges are pushing cells toward the surface to keep pace with the growing hand. These two forces, in addition to the tension created by cell adhesion, cause infolding of the epidermal layers above the attachment site of the secondary ridges [3]. These infoldings are progressively mirrored on the surface as the furrows of friction ridge skin, as secondary ridges continue to mature from 16 to 24 weeks **EGA** [36].

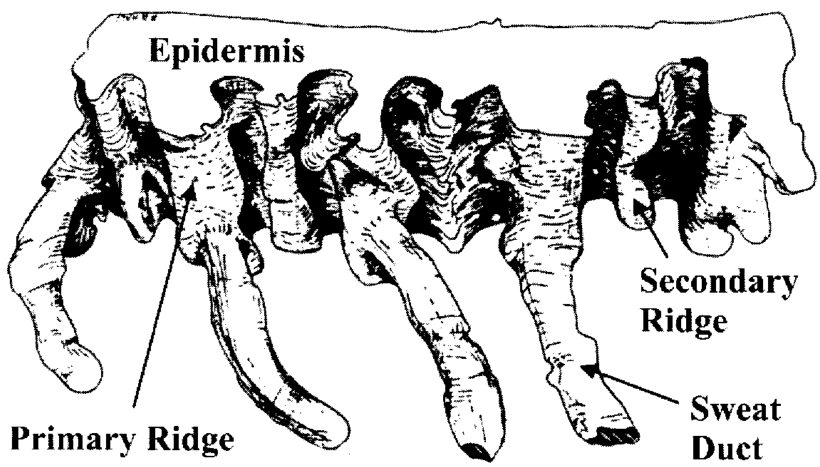


Figure 8

Reconstruction of the underside of fetal volar epidermis, displaying primary ridges with sweat duct formations and the beginning of secondary ridge formation [3].

Incipient Ridges

Little is known about the morphogenesis of incipient (nascent, interstitial, rudimentary) ridges, but several theories could account for their presence in volar skin. Because primary and secondary ridge formation are separate timing events, incipient ridges could be a result of an abnormal transition from primary ridge formation to secondary ridge formation at about 16 weeks **EGA.** One plausible mechanism of incipient ridge formation would involve a small period of time between primary and secondary ridge formation when no new primary ridges are forming. Such a window of time would allow the existing primary ridges to mature slightly before secondary ridges form, locking in the detail. Under these circumstances, incipient ridges could simply be the result of a malfunction in whatever mechanism signals the cessation of new primary ridge formation. Those primary ridges that are in the earliest stages of development when secondary ridge formation begins would become incipient ridges. Another plausible mechanism of incipient ridge formation would involve a malfunction in the timing of secondary ridge formation, causing it to begin twice. The original secondary ridges would then have a secondary ridge on each side of them, causing their structure to become that of a small primary ridge. Either of these mechanisms could account for the presence of incipient ridges as an abnormal timing event during formation that would manifest itself across the entire volar area under development at that time. Further, both mechanisms are consistent with the observation that incipient ridges are inherited [37], as timing abnormalities could very well be based in genetics. Regardless of the mechanism of incipient ridge formation, they are based in the same structure as the surrounding friction ridges, and, therefore, inherit the principles of permanence and individuality that allow for their use in the identification process [38, 39]. Some examiners may be reluctant to rely on incipient ridges when comparing two prints due to absence of some detail in one of the exemplars. The two-dimensional representation of incipient ridges is more susceptible to being affected by deposition pressure due to their size difference and location between two larger structures.

Maturation Process

After maturation of the primary and secondary ridges at 24 weeks **EGA,** anastomoses begin to cross through the dermis [3], linking primary and secondary ridges and molding the upper portion of the dermis into papillae pegs (Figures 9 and 10).

As the skin progresses through this entire process of formation (Figure 11), a nearly infinite number of factors contribute to the end result: complete biological uniqueness, from ridge path down to the structure and shape of a single ridge and beyond. To say that duplication of the entire process of biological formation could occur in any given piece of skin and be indistinguishable from another piece of skin would be equivalent to an identical dump truck load of sticks being scattered twice along the same stretch of road, and saying each stick could land in the exact same position both times.

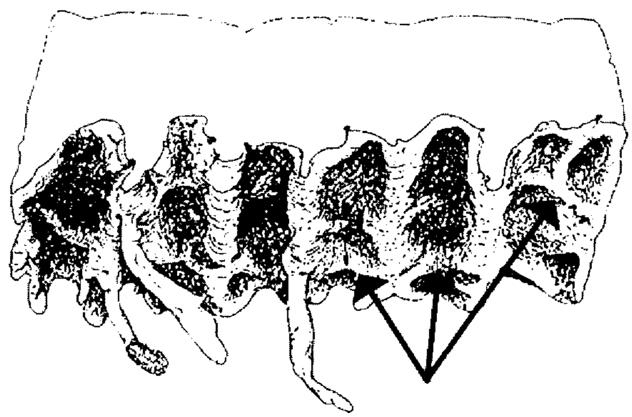


Figure 9

*Reconstruction of the underside of the epidermis of fetal volar skin, the arrows demonstrating sections of epidermis which have bridged primary and secondary ridges, cordoning off sections of dermis commonly referred to as "papillae **pegs**" [3].*

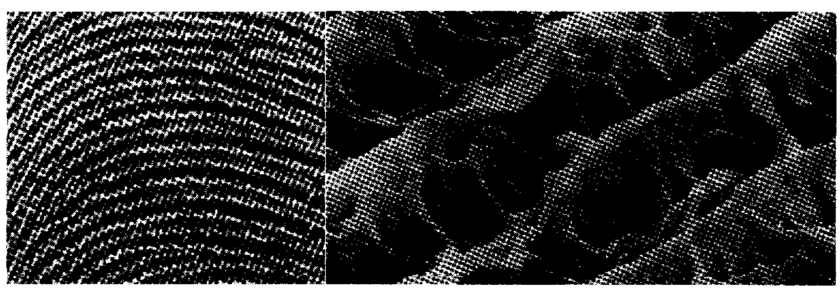


Figure 10

SEM image of the complex undersurface of the epidermis (approximate magnifications left: 8x and right: 80x) [11].

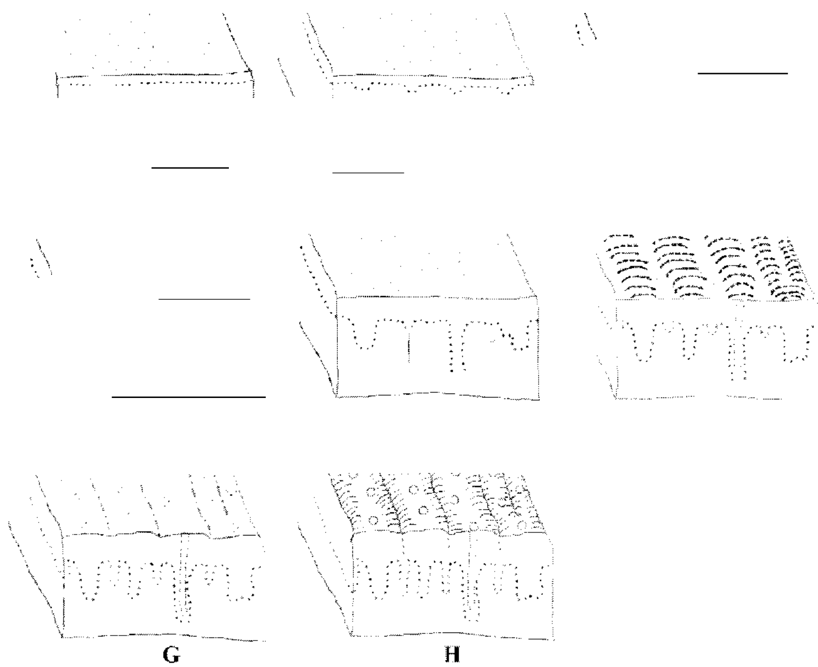


Figure 11

This series qfthree-dimensional drawings represents the epidermul-dermaljunction and the surface ofthe skin before ***(A),*** *during (B-E), and after (F-H)* ***the*** *critical stage of friction skin ,formation.* ***A:*** *The epidermis remains undifferentiated until about 10-11 weeks EGA. B: Primary ridge formation at the epidermal-dermal border, protruding into the dermis. C: Primary ridges continue to penetrate the dermis. D: The skin grows during the critical stage, separating existing primary ridges. E: New primary ridges are thought to form, pulling away from and between existing ridges. F: Secondary ridges begin forming between primary ridges at around 16 weeks EGA, and new primary ridge formation ceases. Surface ridges begin to take form as secondary ridges proliferute downward. Sweat gland ducts have begun to form by this time. G: Secondary ridges continue to mature and surface ridges continue to form. H: By about 24 weeks* ***EG.4,*** *the secondary ridges are upproaching the depth qfthe primary ridges, and the entire system begins the maturation process.*

Pattern Formation

It is seen throughout the physical world that ridges tend to align perpendicularly to physical stress across a surface (Figure 12). Ridges also form transversely to the lines of growth stress in friction skin. The predominate growth of the hand is longitudinal (lengthwise). subsequently. ridges typically cover the volar surface transversely (side to side), as seen in the ridge flow in the joints of the fingers. However, localized eminencies on the volar surfaces of the hands and feet create stresses in directions other than longitudinal, and, therefore, redirect the flow of the ridges in a complex manner across these three-dimensional structures.

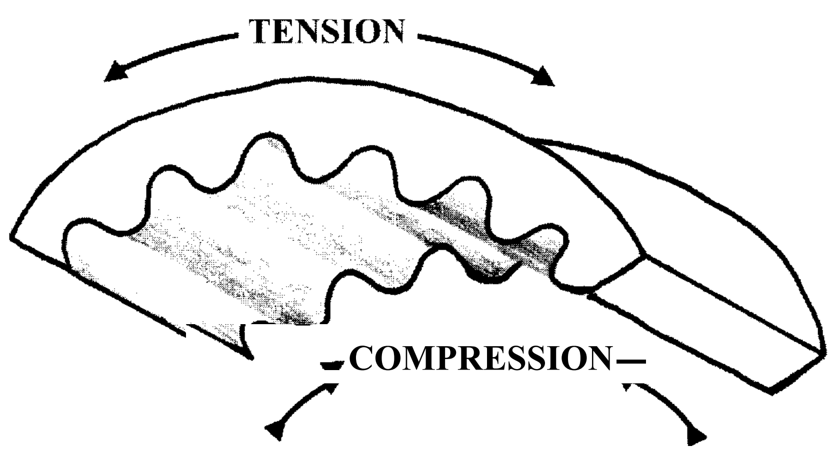


Figure 12

When a semi-flexible membrane isflexed, ridgesform transversely to the forces of tension (on the top) and compression (on the bottom).

Development of the Hand and the Volar Pads

The hand changes topography greatly during the initial phases of formation (Figure 13). At approximately 5 to 6 weeks EGA, the hand is a flat, plate-like structure with thickenings of tissue that show the contours of what will become fingers. From 6 to 7 weeks EGA, these thickenings begin to form the bone models (cartilage at this stage) and muscular components of the hand. Also during this time, the fingers begin to separate and the first volar pads appear on the palm. Volar pads (Figure 14) are transient swellings of mesenchymal tissue under the epidermis (this tissue later forms the dermis) on the palmar surface of the hands and soles of the feet of the human fetus (Figure 15). The interdigital pads appear first, around 6 weeks EGA, followed closely in time by the thenar and hypothenar pads. At approximately 7 to 8 weeks EGA, the volar pads begin to develop on the fingertips, starting with the thumb and progressing toward the little finger in the same radio-ulnar gradient that ridge formation will follow. By 8 weeks EGA, the bone models begin to ossify and the joints begin to form between the bones of the hand. By 8.5 weeks EGA, the hand has an external morphology similar in proportion to the infant [29]. The pads remain well rounded during their rapid growth until about 9 to 10 weeks EGA, after which time they begin to demonstrate some individual variation in both shape and position. These timed events represent the general consensus of the researchers actually observing fetal tissue [4, 33, 34, 40].

As a result of their slowing growth, the pads become progressively less distinct contours on the more rapidly growing hand (Figure 16). This process is referred to as "regression," but it is important to understand that the pad is not actually shrinking. The volar pads of the palm begin to regress first, followed by the volar pads of the fingers. By 16 weeks EGA, volar pads have completely merged with the contours of the fingers, palms, and soles of the feet [29].

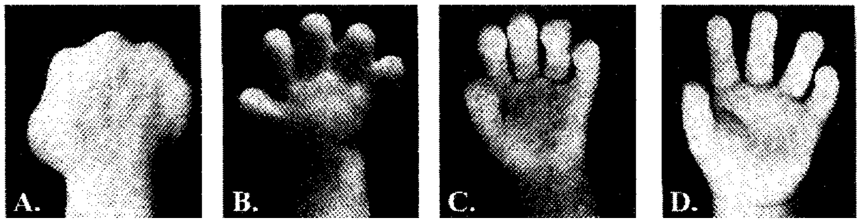


Figure 13

Growth of the hand begins from a paddle-like form (A), continues as the fingers take shape (B), the volar pads form (C), and continues to mature (D) [40].

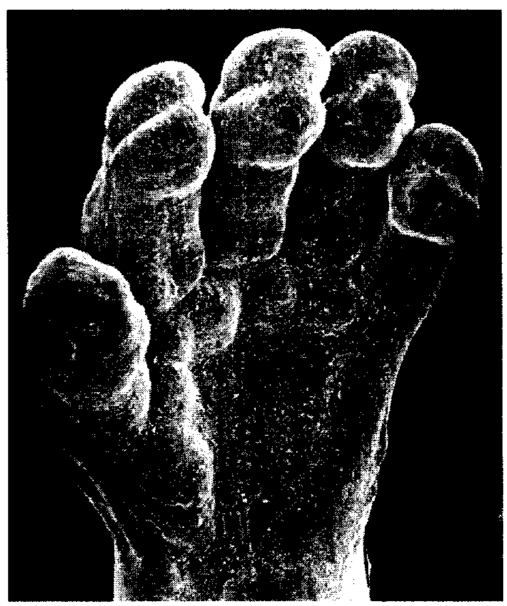


Figure 14

SEM view of the hand of a fetus, displaying prominent volar pads [65].

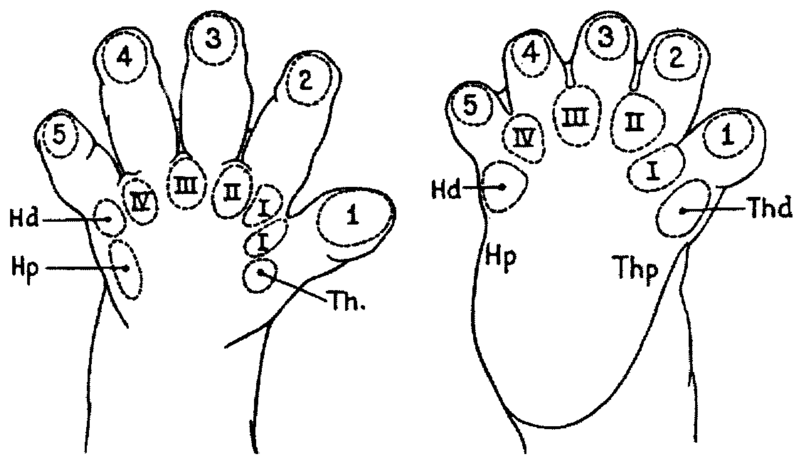


Figure 15

There are normally 11 volar pads on each limb (one on each digit and six on the palmar and plantar surface.) In Cummins' diagram above, he has noted that the hypothenar pad qf the palm is divided into a distal (Hd) and proximal (Hp) portion, and that the first (I) interdigital volar pad (associated with the thumb) is also divided into two portions, making 13 distinct elevations on euch palmar surface. On plantar surfaces, the proximal portions of the hypothenar pad (Hp) and the thenar pad (Thp) are absent, leaving 11 distinct plantar elevations [40].

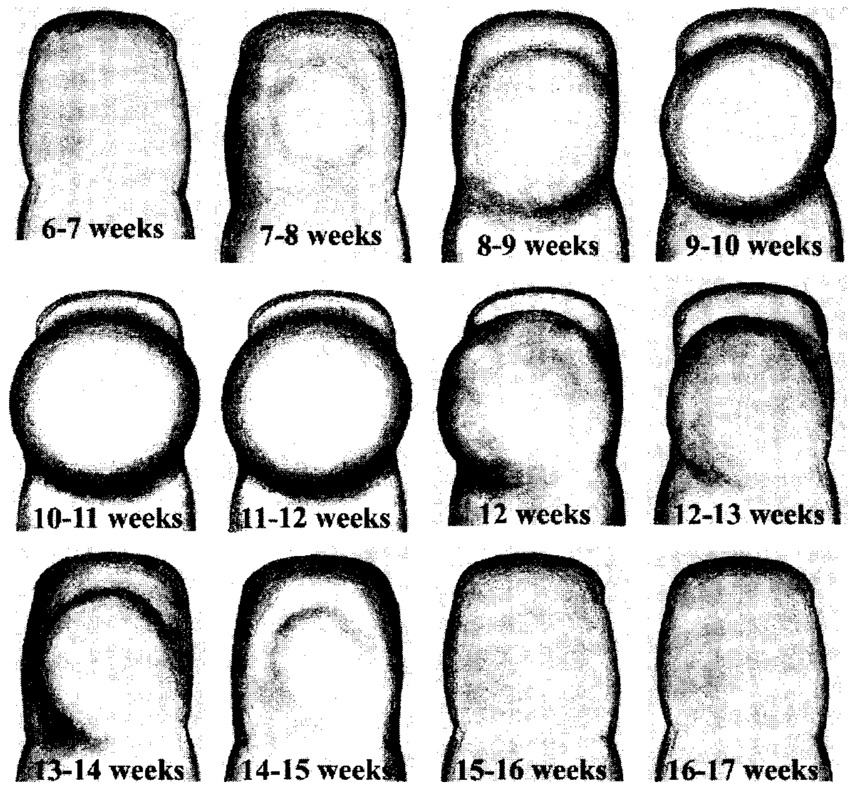


Figure I6

The life of a volar pad fromformation (7-8 weeks EGA on thefingers) until complete regression (I4 weeks EGA). It is understood that all EGA values related to volar pad size and shape are highly variablefromfetus tofetus, and, therefore, are only included as approximations in this general developmental scheme.

Ridge Alignment: Growth Stresses on the Volar Surface

The growth and regression of the volar pads produce variable physical stresses across the volar surface that affect the alignment of the ridges as they first begin to form. Generally, volar pads are high and round when the cells along future primary ridges first begin to proliferate. Ridges form concentrically around the apex of a high round pad, conforming to the navigational pattern of the loxodrome [41] (Figure 17). Research in both the medical and mathematical fields support this same physical model applying across the entire volar surface of the hands and feet [28, 33, 40, 42]. Researchers have observed ridges forming on high, pronounced volar pads conforming to the surface as large-count whorl patterns. Conversely, ridges forming on a finger with a low or absent volar pad are low-count or arch type patterns [34]. The results of this physical model become extremely complex on asymmetrical variations of the spherical shape.

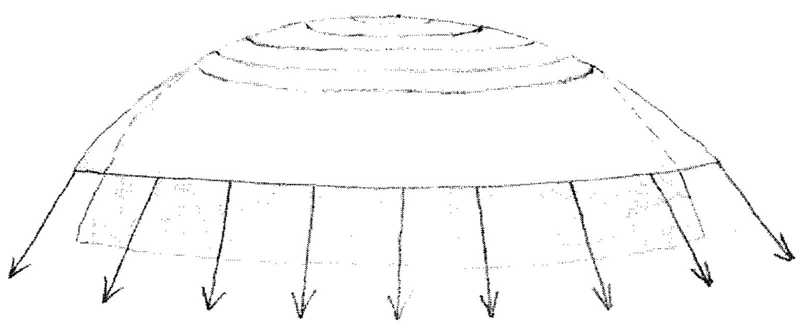


Figure 17

In navigation, this pattern is known us a loxodrome, and resultsfrom an elasticfilm being stretched over a hemisphere. Ridges,form concentrically around the apex. The mathematical formula for this pattern can be found in tensor calculus, afield which offers much promise in predicting what ridgeformations might form across different shaped surfaces.

The effect of the size and shape of the volar pad on ridge patterns has been studied intensely by several researchers over the years. Bonnevie first hypothesized that volar pad height affected fingerprint patterns in 1924 [18]. Three years later, Harold Cummins published an extensive analysis of malformed hands to elucidate the effect of the growth and topology of the hand on ridge direction [33]. Cummins concluded that ridge direction is established by the contours of the hands and feet at the time of ridge formation. Penrose examined fingerprint pattern formation from a mathematical perspective [28, 43].

The distinction between the size. height, and shape of the volar pad, and the effects of differences in each of these elements on a fingerprint pattern is a difficult topic to study, a fact that many recent researchers have acknowledged [44, 45, 46]. Naturally, such differences cannot be evaluated in one individual. Large bodies of data must be studied and correlated in order to deduce which factors may affect specific pattern elements. Fortunately, a literary review has allowed such study. and a unique picture of fingerprint pattern formation has emerged which combines portions of many theories, but goes beyond what has been previously published on this subject.

Ridge Count: Timed Events and Volar Pad Size

The size. particularly the height, of the volar pad during primary ridge formation affects the core to delta ridge count of normal fingerprint patterns [18, 22, 41]. Holt reported that the Total Finger Ridge Count (TFRC) of all 10 fingers is the most inheritable feature in dermatoglyphics [47]. This combined information points directly toward timing events related to volar pad and friction ridge formation affecting fingerprint patterns.

The ridge count ofa fingerprint pattern is related to two different timed events: the timing of the onset of volar pad regression versus the timing of the onset of primary ridge formation. Differences in either timed event will affect the ridge count of that particular pattern if the events are timed independently. For example, early onset of volar pad regression would lead to a volar pad which was in a more regressed state by the time of the onset of primary ridge formation, and a lower ridge count pattern (or arch) would result. On the other hand, overall late onset of volar pad regression would mean that the pad was still relatively large when primary ridges began forming, and, therefore, a high ridge count pattern would result (Figure 18A). This theory is supported by a study which found that "late maturers" had higher than average ridge counts, and "early maturers" had lower than average ridge counts [48]. If volar pad regression onset occurred at the normal time, then earlier than average onset of primary ridge formation would occur on a larger than average size volar pad. This circumstance would lead to a higher than average ridge count. Likewise, later than average onset of primary ridge formation would occur on a smaller than average volar pad. This circumstance would lead to a lower than average ridge count (Figure 18A). When both early and late timing of both factors are taken into account, the results become even more complex (Figure 18B).

Ridge Count: Converging Ridge Fields

The onset of cellular proliferation which begins primary ridge formation occurs first in three distinct areas: (1) the apex of the volar pad (which corresponds with the core of the fingerprint pattern), (2) the distal periphery, or tip of the finger (near the nail bed), and (3) the distal interphalangeal flexion crease area, below the delta(s) in a fingerprint. As ridge formation continues, new proliferation occurs on the edges of the existing ridge fields, in areas that do not yet display primary ridge formation. These three "fields" of ridges converge as they form, meeting in the delta area of the finger (Figure 19). This wave-like process of three converging fields allows for the visualization of how deltas most likely form (Figure 20).

The concept of "converging ridge fields" also offers a way to visualize the difference between the formation of large versus small ridge count patterns. If ridges begin formation on the apex (center) of the pad first and proceed outward before formation begins on the tip and joint areas, then by the time the fields meet, a relatively large distance will have been traveled by the field on the apex of the pad; and a large count pattern would be formed (Figure 21). However, if the ridges form first on the two outermost portions and proceed inward, and formation begins at the last instant on the apex of the pad, then only a few ridges may be formed by the time the fields meet; a very small pattern is observed (Figure 22). The combined observations of different researchers examining fingerprints during the critical stage of development further support the validity of this model [19, 20, 27, 29].

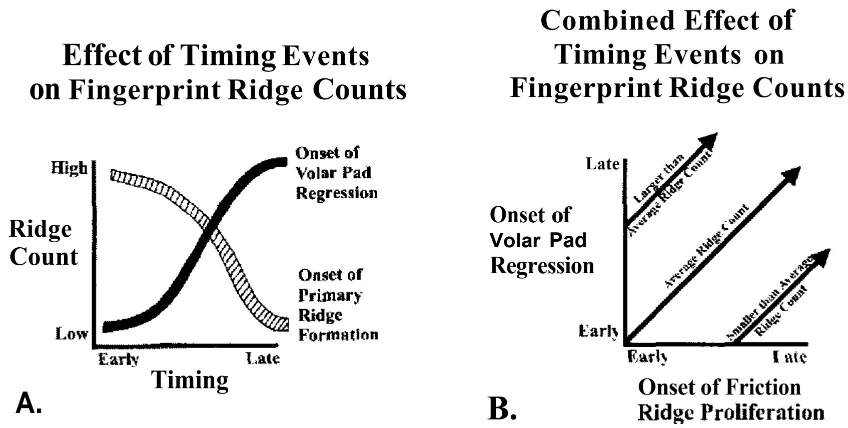
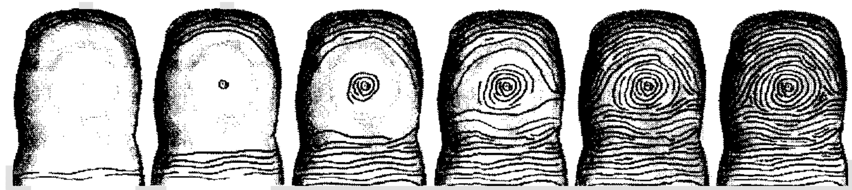


Figure 18

Chart A demonstrates the effects of two independent timing events on the resulting ridge count of a fingerprint pattern. Chart B demonstrates their combined effects on fingerprint pattern ridge count.



Figtire 19

Ridges-form in three distinct locutions on the end joint of thefinger, and converge.

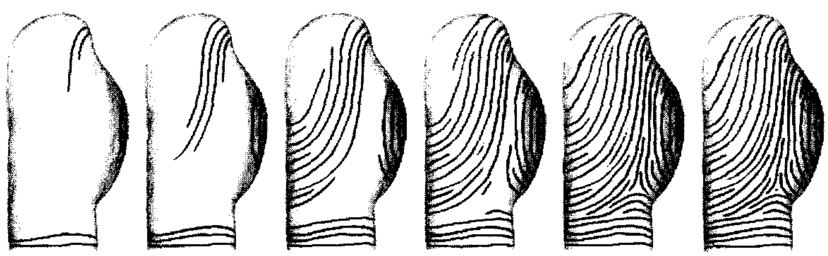


Figure 20

Deltas form where three ridgefields meet.

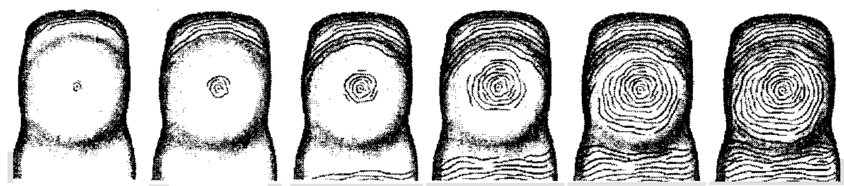


Figure 21

Large count pattern formation: ridges form **in** *the center first and proceed outward before being met by the other ridgefields.*

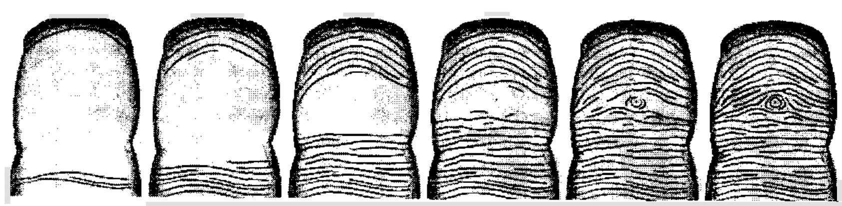


Figure 22

Small count pattern formation: ridges form on *the outer perimeter* of *the pattern area and proceed inward until the last instant, when ridge formation begins ut the apex of the pud.*

Ridge Counts: External Factors

To make matters even more complex, the size of the volar pad with respect to the finger is also affected by many factors. Diet and chemical intake of the mother (IO). hormone levels (44), radiation levels (49), and any other factor that could affect the growth rate of the fetus during the critical stage could all indirectly affect the ridge counts of the developing fingerprints. It is important to remember that anything that could affect the tension across the surface of the finger could affect the resulting ridge count. However. Holt's 1968 findings seem to indicate that timing events, rather than environmental factors, play the dominant role in determining TFRC. **A** significant point to remember is that ridge counts are affected primarily by the combined timing events of volar pad regression and primary ridge formation. Pattern type, on the other hand, is affected by a completely different set of factors.

Pattern Type: Volar Pad Shape and Symmetry

The overall shape and symmetry of the finger volar pad when ridges first begin to form determines pattern type. Cummins, Penrose, and others have long reported that high and round (and/or narrow) volar pads form a large whorl-type pattern, asymmetrical "leaning" pads form looping patterns, and low or absent volar pads form arch patterns [33, 43]. In 1987, Babler validated the correlation between pad symmetry and pattern type [34].

Whether ridge flow will conform to a whorl or a loop pattern depends entirely on the symmetry of the stress across the surface of the finger. If the volar pad and other elements of finger growth are perfectly symmetrical during the onset of primary ridge formation, then a symmetrical pattern (a whorl or an arch) will result. However, if the volar pad and/or other growth factors of the finger are asymmetrical during the critical stage, then that same degree of asymmetry will be reflected in the ridge flow of the resulting pattern. This biological process cannot be thought of as limited to the extremes of regression occurring either totally symmetrical or leaning all the way to one side (totally askew). In fact, there is a continuum involved from whorl patterns to loop patterns. Figure 23 illustrates several patterns from different individuals whose volar pads were roughly the same size at the critical stage (similar ridge counts), but differed in their degree of symmetry. Subtle variations in the symmetry of a volar pad could affect the formation of a whorl pattern versus a central pocket loop whorl pattern, or a central pocket loop whorl pattern versus a loop pattern. Any one of the numerous genetic or environmental factors present during the critical stage could cause a slight deviation in the normal developmental symmetry of the volar pad, and, therefore, affect the resulting pattern type.

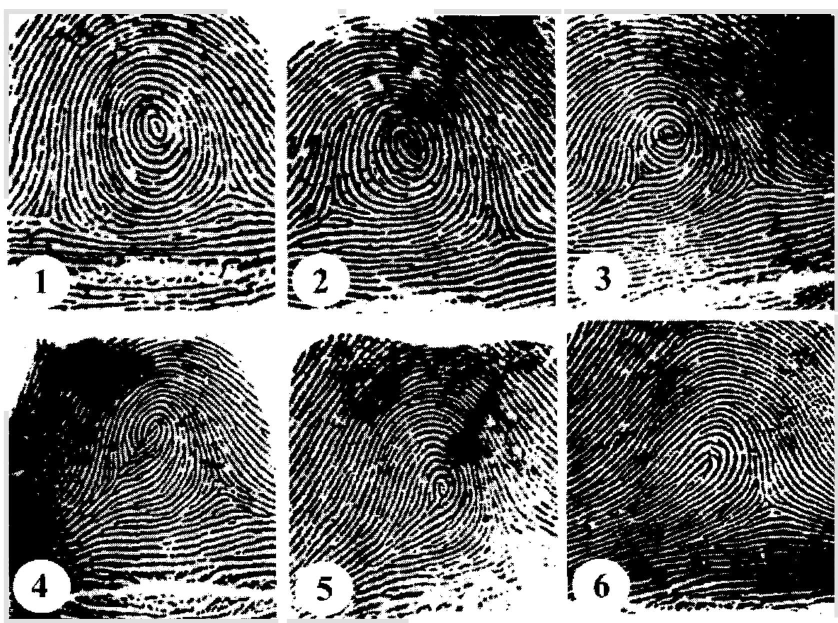


Figure 23

Patterns from different individuals representing the continuum of differing volar pad symmetry. (1) At the onset of friction ridge proliferation, the volar pad of this first pattern was nearly symmetrical. (2) The volar pad of the second pattern was only slightlv displaced, the third slightly more, etc. (6) The volar pad of this pattern was completely displuced.

Timing (Size) and Symmetry of Volar Pads: Cumulative Effect

When it is understood that timing and symmetry control two very different elements of ridge flow, it becomes easy to see how both small and large loop and whorl patterns form. A finger pad that regresses symmetrically will form a whorl pattern, regardless of early or late timing of friction ridge formation with respect to volar pad regression. If the timing of the onset of primary ridge formation in this situation is early in fetal life, then the volar pad will still be high on the finger and the whorl pattern will have a large ridge count. If timing is later in fetal life after the pad has almost completely been absorbed into the contours of the finger. then a low-count whorl pattern will result. Any further regression, and an arch pattern would form (Figure 24). Likewise, asymmetrical finger pads will form loop patterns, and will also be affected by timing. If ridges begin forming early with respect to volar pad regression on an asymmetrical pad, then the pad will be large and a large count loop will result. Later timing leads to a low-count loop or arch-type pattern (Figure 25). Again, it is emphasized that volar pads are not simply symmetrical or asymmetrical, rather, a continuum of symmetry accounts for the variety of pattern types observed.

A regression scheme seems to exist whereby the volar pad is symmetrical at the onset, and becomes progressively asymmetrical as the volar pad regresses. This is supported by general fingerprint pattern statistics which show that over half of all fingerprint patterns are ulnar loops. More specifically, this scheme is supported by fetal research which has determined that early timing of primary ridge formation leads to a higher percentage (95%) of whorls [50]. Also, low and high ridge count patterns occur less frequently than average count patterns [51]. All the data studied tends to indicate that volar pads regress from an early symmetrical position to an asymmetrical position later in fetal life. Although this is the norm, it is certainly not without exception, because whorl patterns with extremely low ridge counts and loop patterns with extremely high ridge counts can both be found with relative ease in large inked print collections.

Figure 24

Thesepatterns were formed on *completely symmetrical volar pads. Top row illustratesfetal condition of volar pad, while bottom row illustrates resultingprint. From left to right the images show the timing offriction ridge proliferation versus volarpud regression. Left: Ridge prolijeration was early, and, therefore, occurred* on *a large volar pad. Right: Ridge proliferation was lute,* und, *therefore, occurred on a small or non-existent volarpad.*

Figure 25

Thesepatterns wereformed on *completely asymmetrical volar pads.Top row illustrates fetal condition of volar pad. while bottom row illustrates resultingprint. From left to right the images show the timing of friction ridge proliferation versus volarpad regression. Left: Ridge prolijeration was early, and, therefore, occurred on a large volarpad. Right: Ridge proliferation was late, and, therefore, occurred* on *a small or non-existent volarpad.*

Unusal Fingerprint Patterns

There are basically two fundamental factors that serve as a starting point during the critical stage of development in determining what pattern will result on a particular finger. The first is the timing of the onset of friction ridge formation with respect to the stage of volar pad regression. The second factor is the degree of volar pad symmetry (or asymmetry). If these two factors can be evaluated independently, then insight may be gained with respect to causal factors of unique patterning. Tension across the surface does not always conform to normal models, which is why we see double-loop whorls, accidental whorls, nutant loops, cuspal patterns, etc. These abnormal patterns could be caused by irregular volar pad growth or regression, unique growth of the bony distal phalanx, physical pressure on the digit, or any other factor affecting the symmetry of the volar pad. Perhaps a lack of tension across a portion of the pattern is the cause of dissociated ridges, or a complete lack of tension across the entire digit (stagnant growth rate) might explain dysplasia. Perhaps the fundamental reason that individuals with certain chromosomal diseases, such as Down's syndrome and Turner's syndrome, consistently display abnormal ridge counts [52, 53, 54] is that chromosomal abnormalities are closely linked with critical timing events in the development of the nervous system and other important fetal milestones during the 10.5 to 16 week **EGA** critical stage. The interrelated factors of timing and symmetry seem to be most significant in affecting tension across the surface, and account for the wide range of dermatoglyphic patterns seen by latent print examiners around the world.

Palm Prints and Vestige Patterns

Friction skin formation concepts apply to all volar skin, whether on the finger, palm, or sole. This paper primarily focuses on the fingers, because fingerprint patterns are the most widely discussed and studied in the literature. However, patterning on the palm and sole could be addressed using these same concepts. Volar pads and timing events still affect patterning, but terminology regarding ridge counts and pattern types would have to be redefined for application to each relevant area. The same concepts apply on the surface of the palm with respect to volar pad location and delta (triradii) formation: triradii in the palm also form along the shoulders of the volar pads, which have mostly regressed into the surface of the palm before the critical stage. Occasionally, patterning can be found in the thenar or hypothenar areas, but these volar pads are the first to regress and are usually completely absent during the onset of primary ridge formation. Palm prints open our eyes to the formation of friction skin over the entire volar surface. Congenital lymphedema (swelling of the limbs during fetal formation) is but one example which demonstrates this through both high ridge count whorl patterns on the fingers and complex palmar patterning. However, it can be generally said that the same formation concepts apply across all volar areas: tension during the critical stage conditions ridge alignment, stresses across small areas determine minutia placement, and heterogeneous basal cell distribution determines ridge shape and pore location.

Some flexion creases form concurrently with friction ridge skin [56]. Disruptions in ridge flow are sometimes found around creases in certain areas undergoing simultaneous development. The most common area is the distal transverse flexion crease (top crease) just under the little finger [57]. Ridges in this area occasionally appear to turn abruptly into the crease, suggesting the crease and ridges were both forming during the critical stage. The scope of this paper did not include an in-depth study of volar creases. Suffice it to say that flexion creases are part of the same volar skin structure just as ridges, and. therefore, share the same principles of permanence and individuality [56, 58].

The formation of vestige patterns (Figure 26) has not been addressed in depth in the literature, and is not widely discussed in the field. However, several researchers have independently studied and diagrammed the volar pads on the surface of the palm (Figure 27). When volar pads between the index finger and the thenar area in the palm are pronounced during the critical stage, it can be seen that the surface relief, as depicted by Kimura [56], would lead to the formation of a vestige pattern. (Figure 28)

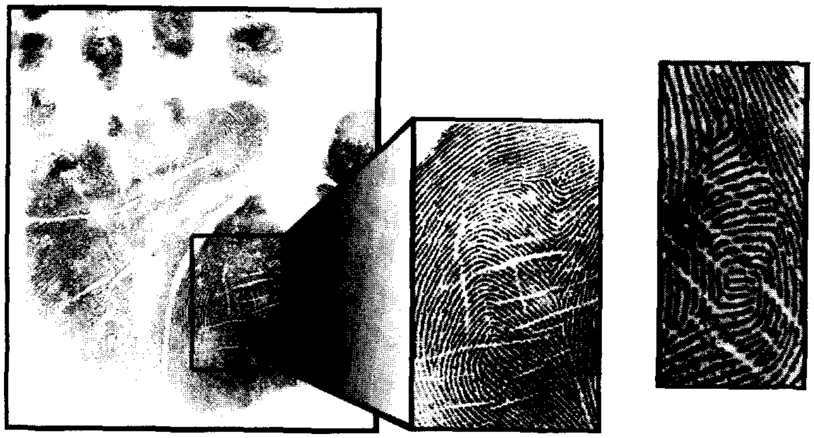


Figure 26

Vestigepatterns can sometimes be found in the thenar area ojpalm prints.

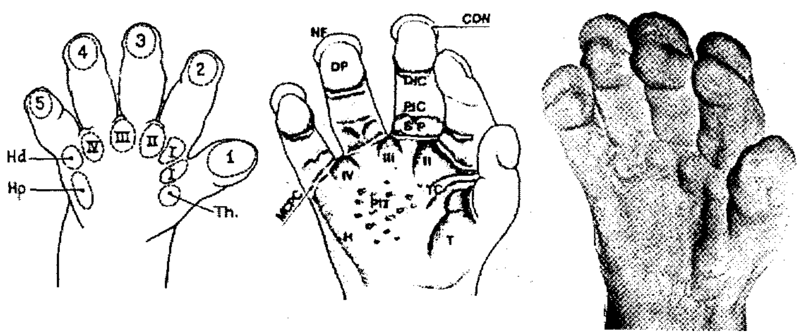


Figure 27

Researchern' observations ure consistent throughout the literature in representing aportion of the first palmar volar pad being separate from the rest, on occasion being pushed up next to the thenurpad [40, 56, 66, respectively].

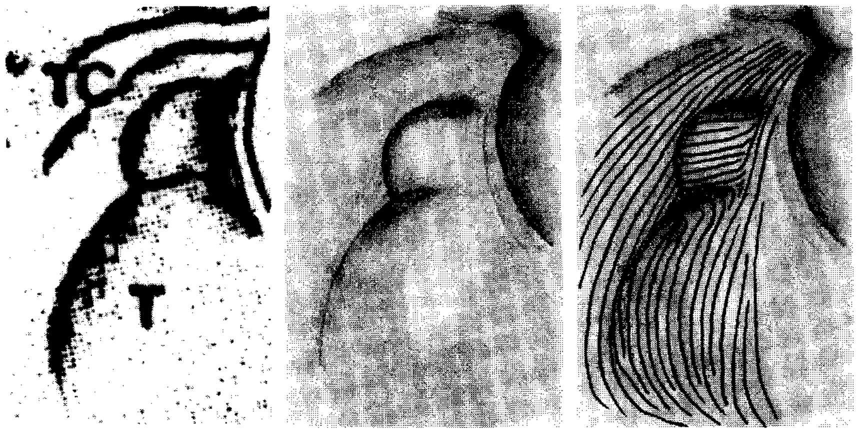


Figure 28

*When the **relief** of the palmar sui-face, as drawn by Kimura, is present, it becomes easy to **see** how vestige patterns likely form.*

Pattern Type: Bone Morphology

As demonstrated by fingerprint pattern frequency statistics, volar pads tend to regress in a manner conducive to the formation of ulnar loops. The cause of radial loops is simply a volar pad that is “leaning” or regressing in the opposite direction. In 1987, Babler found that radial loops and the distinctive ridge flow on the tips of the thumbs are associated with unique shapes in the bony distal phalanx [34]. Babler also reported a higher frequency of whorls associated with shorter distal phalanges and a higher frequency of whorls in instances where bones were less ossified [35]. Unique bone development in the end joint of the finger during the critical stage could affect tension across the skin, which, in turn, affects friction ridge patterning.

The Role of Genetics

Every aspect of the growth and development of a single cell to a fully formed human is directed by genetics. The capacity to form friction ridges is inherent. The patterns that these ridges form, however, are limited by nature and are defined as whorls, loops, arches, combinations and transitions of these basic patterns, or lack of a pattern [17]. Nature has established patterns, while genetics directs when and where ridges will form.

As with all biological traits, genetics does not independently control the resulting patterns on friction ridge skin. The ultimate example is monozygotic twins, who share identical genetic information and very similar environments, but on many occasions have very different patterns. The role of genetics is currently understood by the indication that several main genes, in conjunction with a number of modifying genes, may be responsible for volar patterning, but that patterning is affected by the environment [17, 59, 60, 61]. These genes most likely influence pattern formation indirectly through timing events, volar pad regression, growth rate of the fetus, and other factors which significantly influence pattern type and ridge count. Stresses across small areas of skin are not inherited, and represent but one of many environmental factors which influence pattern formation.

Until recently [45, 46], most researchers in the field of genetics and physical anthropology have traditionally viewed TFRC as evidence of direct genetic control of fingerprint pattern formation [18, 47]. The research of Holt, published in 1968, regarding the heritability of TFRC is a significant finding, and supports the two-tiered development scheme suggested by our study. Genetically controlled timed events would be less susceptible to environmental variations, and, therefore, TFRC would be more inheritable than pattern type. However, fingerprint pattern type and ridge count are ***indirectly*** inherited, and, therefore, are not affected by only one developmental factor. Ridge flow and ridge count are both affected by tension across the surface of growing fetal skin, as visualized through the model of differential timing (Figure 29).

Thousands of anthropological studies [62] have been conducted on distinct populations to identify trends in fingerprint pattern frequencies. Additionally, the medical community has been, and continues to be, very interested in dermatoglyphics as an indicator of abnormal fetal development during the critical stage [63, 64]. Using this new understanding of friction ridge and pattern formation, it may be possible to re-examine some of the massive amounts of previously published data, isolate certain aspects of the fingerprint patterns, and provide even further insight into the mechanism of friction ridge formation.

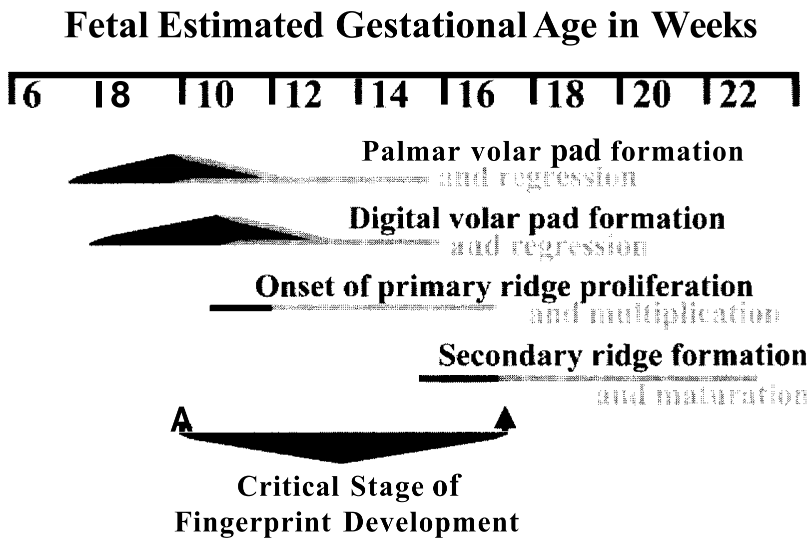


Figure 29

Summary of the timed events critical to the development offingerprints, the exact timing of which varies among individuals. Black represents event onset, gray represents pad regression and ridge maturation.

Conclusion

Although different terminology was found throughout the literature in this study, different elements and general concepts contribute to form an over-all picture of the structure of skin (Figure **30).**

The fundamental principle of permanence is based in the structure of friction ridge skin. The cells of the epidermis rise in concert from the constant but unique production of the basal layer, providing an outer layer that consistently represents the unique cell arrangements and production of the landscape from which it emerged. Within the skin, there are three primary structural elements which enforce the permanence of the friction ridges: (1) the attachment of epidermal cells to each other, (2) the attachment of the basal epidermal cells to the basement membrane, and (3) the attachment of the dermis to the basement membrane. The properties of the basement membrane zone are consistently represented on the surface through the constant supply of new skin cells from that template. In addition to the typical structure of skin, the enhanced structure of friction ridge skin, with its alternating primary and secondary ridges, further anchors the surface ridges and furrows.

The uniqueness of friction skin is imparted to the permanent base structure from a sea of random forces, which, themselves, are affected by a seemingly infinite number of factors. The fetal volar pads play a major role in affecting the tensions that directly influence pattern formation (volar pad symmetry) and ridge count (volar pad size), but minutia formation occurs on a much smaller level. Localized stresses (tensions and compressions), resulting from growth of the tissue layers of the digit and interactions with existing ridge fields, create the foundations for second level uniqueness. Ridge morphology (third level detail) demonstrates a unique heterogeneous cellular community along the basement membrane, which constantly feeds the epidermis a three-dimensional portrait of its collective individuality. It is completely inconceivable that these physical stresses and cellular distributions could be exactly duplicated, on any level, in two different areas of developing fetal tissue. The fact is that each individual ridge is unique. Therefore, any ridge arrangement, regardless of quantity, is unique and could only come from one source. Wide variations in the amount of detail that transfers from the three-dimensional to the two-dimensional realm during any given contact may not permit individualization, but the arrangement itself is still unique.

More than ever, latent print examiners are expected to completely and accurately describe the principles upon which the science of friction skin individualization is founded: the concepts of *permanence* and individuality that permit the use of fingerprints as a means of identification. The scientific basis for stating that friction ridges and ridge formation are permanent and unique can be found in the biological sciences. It remains the responsibility of each expert, in every case worked, to be prepared to address these issues, defend the examination philosophy and methodology, and uphold a century of scientific excellence in the field of latent print examinations.

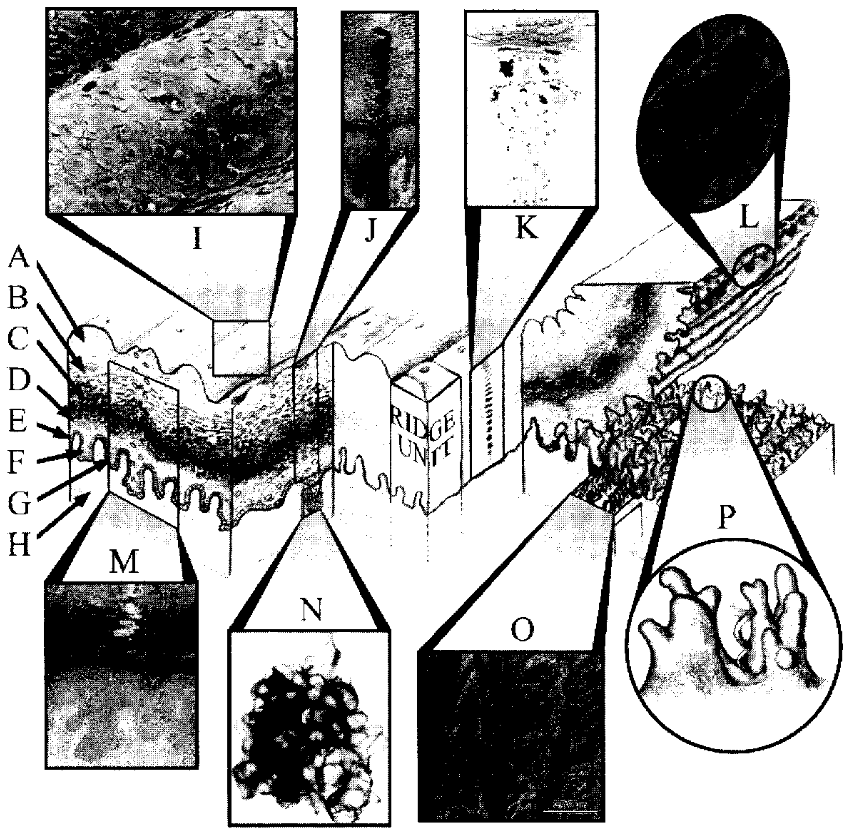


Figure 30

A: Stratum Corneum, or Horny layer
B: Stratum Lucidum, or Hyalin layer
C: Stratum Granulosum, or Granular layer
D: Stratum Spinosum, or Spinous layer
E: Stratum Basal, or Basal (Generating) layer
(A-E): Layers of the Epidermis
F: Primary Ridge (associated with surface ridge)
G: Secondary Ridge (associated with surface furrow)
H: Dermis
I: SEM view of a surface ridge [11]
J: Spiraling sweat duct [11]
K: Epidermal cell progression [11]
L: Under-surface of the epidermis [11]
M: Stained cross-section of volar skin [66]
N: Coiled sweat gland [11]
***O:** Capillaries underneath dermal papilla [66]*
P: Close-up of dermal papilla [5]

Acknowledgements

We would like to thank the organizations and individuals who have supported our efforts and helped us in this venture: The Mississippi Crime Laboratory, Alaska Scientific Crime Detection Laboratory, and the Henderson, Nevada Police Department for affording time and resources to complete this project; William Babler for allowing us to see friction ridge formation through his eyes as he observed from his studies over the years; David Ashbaugh for his valuable insight and experience from years of research; Pat Wertheim, Dave Grieve, and Alan McRoberts for their inspiration and encouragement; and the many researchers from whose studies this paper was made possible.

For further information, please contact:

Kasey Wertheim, CLPE, CSA
Forensic Scientist II
Mississippi Crime Laboratory
Meridian Branch Laboratory
P.O. Box 4450
Meridian. MS 39304-4450

(601) 483-5273
kwertheim@mcl.state.ms.us

Animated illustrations for Figures 6, 7, 16, and 19 - 22 are available at www.clpex.com/animation.htm

References

1. Dauber v. Merrill Dow Pharmaceuticals. United States Supreme court: 509 U.S. 579, 113 **S.** Ct. 2786, 125 L. Ed. 2nd 469. 1993.
2. Holbrook, K. Structure and Development of the Skin. *Pathophysiology of Dermutologic Diseases,* 2nd ed.; Soter, M., Howard, P., Baden, Eds.; McGraw-Hill: New **York,** 1991, pp 3-43.
3. Hale. A. Morphogenesis of Volar Skin in the Human Fetus. ***Am.*** *J. of Anat.* **1952.** 91(1), **pp** 147-173.
4. Babler, W. Quantitative Differences in Morphogenesis of Human Epidermal Ridges. In *Dermatoglyphics - Fifty Years Later,* Birth Defects Original Article Series; March of Dimes: Washington DC, **1979,** 15(6), pp 199-208.
5. Misumi, Y.; Akiyoshi, T. Scanning Electron Microscopic Structure of the Finger Print as Related to the Dermal Surface. *Anat. Rec.* **1984,** *208,* **pp** 49-55.

6. Okajima. M. A Methodological Approach to the Development of Epidermal Ridges Viewed on the Dermal Surface of Fetuses. *Progress in Dermatoglyphic Research*, Alan R. Liss, Inc.: NY. 1982, pp 175-188.

7. *The Science of Fingerprints;* Federal Bureau of Investigation, U.S. Department of Justice, U.S. Government Printing Office: Washington, DC, 1987, p. 127

8. Cummins, H.; Midlo, C. *Fingers, Palms, and Soles*. Dover Publications: NY, 1961. pg 39.

9. Chacko, L.; Vaidya, M. The Dermal Papillae and Ridge Patterns in Human Volar Skin. *Acta Anat.* **1968.** 70(1), pp 99-108.

10. Holbrook. K. A. Structure and Function of the Developing Human Skin. In *Biochemistry and Physiology of the Skin,* Goldsmith, L. A., Ed.; Oxford Univ. Press: New York, 1991, Vol. 1, pp 64-101.

11. Montagna, **W.;** Parakkal, P. *The Structure and Function of Skin,* 3rd ed.; Academic Press: Orlando, Florida, 1974.

12. McRoberts, A. The Union Institute, Los Angeles, CA. *Dermatoglyphics: Anatomy and Morphology of Volar Skin*. Unpublished, 1999, pp 1-25.

13. Ashbaugh. D. *Quantifative-Qualitative Friction Ridge Analysis,* CRC Press: Boca Raton Florida. 1999.

14. Okajima. M. Development of Dermal Ridges in the Fetus. *J. of Med. Genetics* **1975,** 12. pp 243-250.

15. Lavker, R.; Sun, T. Epidermal Stem Cells. *J. of Invest. Derniatologv* **1983.** 81:1. **pp** 121-127.

16. Ball, P. *The Self-Made Tapestry; Pattern Formafion in Nature,* Oxford University Press: NY, 1999, pp 1-15.

17. Hirsch. W. Biological Aspects of Finger Prints, Palms and Soles. *Fingerprint and Identification Magazine* **1964,** pp 3-17.

18. Bonnevie, K. Studies on Papillary Patterns of Human Fingers. *J. of Genetics* **1925,** 15, pp 1-112.

19. Hirsch, W.; Schweichel, J. Morphological Evidence Concerning the Problem of Skin Ridge Formation. *J. of Ment. Defic. Res.* **1973,** 17, pp 58-72.

20. Dell. D.; Munger, B. The Early Embryogenesis of Papillary (Sweat Duct) Ridges in Primate Glabrous Skin: The Dermatotopic Map of Cutaneous Mechanoreceptors and Dermatoglyphics. *J. of Comp. Neurol.* **1986,** 244, pp 511-532.

21. Moore, S.; Munger. B. The Early Ontogeny of the Afferent Nerves and Papillary Ridges in Human Digital Glabrous Skin. *Developmental Brain Research* **1989, 48,** pp 119-141.

22. Siervogel, R.; Roche, A.; Roche, E. Developmental Fields for Digital Dermatoglyphic Traits as Revealed by Multivariate Analysis. *Human Biology* **1978.** 50(4), pp 541-556.

23. Meier. R. Sequential Developmental Components of Digital Dermatoglyphics. *Human Biology* **1981,** 53(4), pp 557-573.

24. Roberts. D. Population Variation in Dermatoglyphics: Field Theory. *Progress in Dermatoglyphic Research,* Alan R. Liss: NY, 1982, pp 79-91.

25. Heimer, L. *The Human Brain and Spinal Cord: Functional Neuroanatomy und Dissection Guide,* 2nd ed.; Springer-Verlag: NY, 1994.

26. Smith, L. T.: Holbrook, K. Embryogenesis of the Dermis in Human Skin. *Pediatric Dermatology* **1986,** 3(4), pp 271-280.

27. Babler, W., Marquette University, Milwaukee. WI. Personal Communication, **October 1999.**

28. Loesch. D. The Contributions of L.S. Penrose to Dermatoglyphics. *J. of Mental Deficiency Research* **1973,** 17(1), pp **1-17.**

29. Babler, W. Embryologic Development of Epidermal Ridges and Their Configurations. In *Dermatoglyphics: Science in Transition,* Birth Defects Original Article Series: March of Dimes: Washington DC, **1991,** 27(2), pp 95-112.

30. Evatt. E. J. The Development and Evolution of the Papillary Ridges and Patterns of the Volar Surface of the Hand *J. of Anat.* **1906.** 41, p 66-70.

31. Ashbaugh, D. Ridgeology. *J. For. Ident.* **1991,** 41(1), pp 16-64.

32. Wilder, H.; Wentworth. B. *Personal Identification,* Gotham Press: Boston. 1918.

33. Cummins, H. Epidermal Ridge Configurations in Developmental Defects, With Particular Reference to the Ontogenetic Factors Which Condition Ridge Direction. *Amer. J. Anat.* **1926.** 38(1), pp 89-151.

34. Babler, W. Prenatal Development of Dermatoglyphic Patterns: Associations with Epidermal Ridge, Volar Pad and Bone Morphology. *Collegium Antropologicum* **1987,** 11(2), pp 297-303.

35. Babler, W. Prenatal Communalities in Epidermal Ridge Development. *Trends in Dermatoglyphic Reseurch,* Kluwer Academic Press: Dordrecht, Netherlands, 1990. pp 54-68.

36. Babler, W. Monitoring Patterns of Prenatal Skeletal Development. In *Derniatoglyphics - Fifty Years Later,* Birth Defects Original Article Series; March of Dimes: Washington DC. **1979,** 15(5A), pp 25-38.

37. Penrose, L.; Plomley, N. Structure of Interstitial Epidermal Ridges. *Z. Morphol Anthropol* **1969.** 61, pp 81-84.

38. Ashbaugh, D. Incipient Ridges and the Clarity Spectrum. *J. For. Ident.* **1992.** 42(2), pp 106-114.

39. Okajima, M. Dermal and Epidermal Structures of the Volar Skin. In *Derniatoglyphics - Fifty Years Later.* Birth Defects Original Article Series; March of Dimes: Washington DC. **1979.** pp 179-198.

40. Cummins, H. The Topographic History of the Volar Pads (walking pads; Tastballen) in the Human Embryo. *Contr. Embryol.* Carnegie Institution of Washington: Washington. DC. **1929,** 20, pp105-126.

41. Mulvihill, J.; Smith, D. The Genesis of Dermatoglyphics. *J. Prdiatrics* **1969,** 75(4), pp 579-589.

42. Penrose, L.; O'Hara, P. The Development of Epidermal Ridges. *J. Med. Genetics* **1973,** 10, **pp** 201-208.

43. Penrose, L. Dermatoglyphics. *Sci.* ***Amer.*** **Dec. 1969,** pp 72-84.

44. Jamison, **C.** Dermatoglyphics and the Geschwind Hypothesis I. Theoretical Background and Palmar Results of Dyslexia 11. Digital Results of Dyslexia and Developmental Implications. In *Trends in Dermatoglyphic Research,* Durham, N. M.; Plato, C. C., Eds.; Kluwer Academic Publishers: 1990, pp 99-135.

45. Mavalwala, J.; Mavalwala, P.: Kamali, S. Issues of Sampling and of Methodologies in Dermatoglyphics. In *Dermatoglyphics. Science in Transition,* Birth Defects Original Article Series; March of Dimes: Washington DC. **1991.** 27(2), pp 291-303.

46. Chakraborty. R. The Role of Heredity and Environment on Dermatoglyphic Traits. In *Dermatoglyphics: Science in Transition.* Birth Defects Original Article Series; March of Dimcs: Washington DC. **1991.** 27(2), pp 151-191.

47. Holt, S. *The Generics of Dermal Ridges,* Charles C Thomas: Illinois, 1968.

48. Meier, R.: Goodson, C.; Roche, E. Dermatoglyphic Development and Timing of Maturation. *Human Biology* **1987.** 59(2), pp 357-373.

49. Bhasin, M. Effect of Natural Background Radiation on Dermatoglyphic Traits. *Acta. Anthropogenet* **1980,** 4, pp 1-27.

50. Babler, W. Prenatal Sclection and Dermatoglyphic Patterns. *Amer.n J. Phys. Anthr.* **1978.** 48, pp 21-28

51. Cowger. J. *Friction Ridge* ***Skin.*** Elsevier Science Publishing: NY, 1982.

52. Reed. T. Impact of Changes in Medical Genetics on Teaching and Disseminating Information on Dermatoglyphics. In *Dermatoglyphics: Science in Transition*, Birth Defects Original Article Series; March of Dimes: Washington DC. **1991.** 27(2), pp 305-319.

53. Miller. J. Dermatoglyphics. Birth Defects Original Article Series; March of Dimes: Washington DC. **1981,** 17(2), pp 67-78.

54. Netley, C.; Rovet, J. Verbal Deficits in Children with 47. XXY and 47, XXX Karyotypes: A Descriptive and Experimental Study. *Brain and Language* **1982,** 17, pp 58-72.

55. Schaumann, B.; Opitz, J. Clinical Aspects of Dermatoglyphics. Birth Defects Original Article Series: March of Dimes: Washington DC, **1991,** 27(2), pp 193-228.

56. Kimura, S. Embryological Development of Flexion Creases. Birth Defects Original Article Series; March of Dimes: Washington DC, **1991**, 27(2),pp 113-129.

57. Smith. R. Meridian, MS. Palm Print Symposium. Unpublished, **1998.**

58. Ashbaugh, D. Palmar Flexion Crease Identification. *J. For. Ident.* **1991,** 41(4), pp 255-273.

59. Slatis, H.: Katznelson. M.; Bonne-Tamir, B. The Inheritance of Fingerprint Patterns. *Amer. J. Human Genetics* 1976, 28. pp 280-289.

60. Loesch, D. *Quantifative Dermatoglyphics: Classification, Genetics, and Pathology,* Oxford University Press: NY. 1983.

61. Weninger, M.; Aue-Hauser, G.; Scheiber, V. Total Finger Ridge-Count and the Polygenic Hypothesis: A Critique. *Human Biol.* **1976,** 48(4), pp 713-725

62. Mavalwala, J. *Dermatoglyphics, An Infernational Bibliography,* Mouton Publishers: Chicago, 1977.

63. Kahn, H.; Ravindranath, R.; Valdez, R.; Venkat Narayan, K. Fingerprint Ridge-Count Difference between Adjacent Fingertips (dR45) Predicts Upper-Body Distribution: Evidence for Early Gestational Programming. *Amer. J. Epidemiology,* **2001,** 153(4), pp 338-344.

63. *The State of Dermatoglyphics: The Science of Finger and Palm Prints,* Durham, N., Fox, K.. Plato, C., Eds.; Edwin Mellen Press: New York, 2000.

65. Carlson, B. *Human Embryology and Developmental Biology.* Mosby: NY, 1999. p. 152.

66. Montagna. W.; Kligman, A.; Karlisle, L. *Atlas of Normal Human Skin,* Springer-Verlag: NY, 1992.

Copyright Permission

Figure 2a	Reprinted by permission of March of Dimes Archives Office
Figure 2b	Reprinted from U.S. Government
Figure 3	Reprinted by permission of Harcourt, Inc.

Meagher, Stephen, Bruce Budowle, and Donald Ziesig. 50K vs. 50K Fingerprint Comparison Test: Government Exhibits 6-8 and 6-9 in Pretrial Hearing, *United States v. Mitchell*, Cr. No. 96-407, E.D. Pa. (1999). Document in public domain.

50K vs. 50K Fingerprint Comparison Test

At the request of the FBI, Lockheed Martin conducted a pair of experiments to determine the similarity of minutia patterns between all members of a set of fifty thousand (50,000) fingerprints extracted from the Criminal Master File (CMF).

The fingerprints were restricted to the Pattern Class "Left Loop" of White Males. This restriction increased the likelihood of the fingerprints being similar to one another since the general flow of the ridges for all fingerprints in the set is the same. The number of fingerprints was limited to 50,000 so that the experiments would be completed in time for hearing preparation. The 50,000 "White Male Left Loops" were accumulated from the CMF with no other ordering - the first 50,000 Left Loop records that were chosen by the database retrieval software from the CMF were used for the experiments.

EXPERIMENT 1 DESCRIPTION:

The first experiment compared the minutia pattern of each of the fingerprints in the set with itself and with the minutia patterns of all other fingerprints in the set.

Experiment 1 required two billion, five hundred million (2,500,000,000) comparisons.

Each comparison was performed by two totally different software packages, developed in two different countries by two different contractors using independent teams of fingerprint and software experts. The results of both comparisons were mathematically "fused" using software developed by a third contractor.

The comparison software is:

1 Attributed Relational Graph matcher developed in the USA by Lockheed Martin, Inc.
2 Super Match II matcher developed in France by Sagem.
3 Statistical fusion software developed in the USA by CALSPAN, Inc.

The two "matcher" programs calculate a measure of similarity between the minutia patterns of two fingerprints. In both cases, the scores of an identical mate fingerprint is normalized to 1.0 (or 100%). The statistical fusion program combines the two scores by analyzing the most similar 500 (out of 50,000) minutiae patterns. The fusion operation discards 49,500 very dissimilar minutia patterns before calculating the fusion statistics. As in the case of the "matcher" programs, the fused similarity measure calculated by the fusion program is normalized to 1.0 (or 100%).

We produced the graph titled "Minutia Pattern Similarity (Rolled Prints vs. Rolled Prints)" showing the similarity of the second most similar minutia pattern for all three operations for all fifty thousand fingerprints in the test set. Note that the two "matcher" programs use very different mathematics so it is not unexpected that the distribution of the similarity measures appears different in the chart. This chart shows the operation of the matchers and fusion software on the 50,000 test set.

Next we determined the probability that the degree of similarity of the mate with itself could occur by chance, given the set of 500 most similar fingerprints. This was done using a statistical measure of separability known as "Sample Z Score". (The mathematical basis for this operation is described in "Statistics, Seventh Edition" - McClave, James T., et. al., Prentice Hall, ISBN 0-13-471542) The Z score can be converted into a probability using standard mathematics functions. (The result of these functions are tabulated in "Handbook of Mathematical Functions" - Abramowitz, Milton and Stegun, Irene A., Dover Publications. Library of Congress Catalog Card Number: 65-12253).

Significant Values of Z scores are:

Z	Probability	Explanation
0.0	1/2	Half the scores of a random sample will be higher (or lower).
3.0	1/500	Highest expected score in a random population of 500.
3.4	1/3000	
6.9	$1/10^{11}$	1 chance in 1 followed by 11 zeroes.
21.0	$1/10^{97}$	1 chance in 1 followed by 97 zeroes.

EXPERIMENT 1 RESULTS:

In the 50,000 vs. 50,000 rolled print similarity comparison experiment, we discovered three fingerprints whose non-mate Z scores were greater than 3.0. These were unlikely to occur and were a cause for concern:

Mate Index	Non-mate index	Z Score	Probability (approximate)
48541	48543	6.98	1/100,000,000,000
48543	48541	6.95	1/100,000,000,000
18372	18373	3.41	1/3,000

In all three cases, we found that the non-mate fingerprint had the same FBI Identification number as did the mate:

Index	FBI Identification Number
48541	[redacted]WA1
48543	[redacted]WA1
18372	[redacted]WA9
18373	[redacted]WA9

All three consolidations (high non-mate Z scores) were the result of different rollings of the same fingers.

Aside from these four fingerprints, the best (largest) non-mate Z-score was 1.83 for a probability of approximately 1/20 which is well within the 1/500 expected for 500 random samples.

Similarly, the worst mate Z score was 21.7, for a probability of approximately $1/10^{97}$ (1 followed by 97 zeros).

EXPERIMENT 1 CONCLUSIONS:

The probability of a non-mate rolled fingerprint being identical to any particular fingerprint is less than $1/10^{97}$ (1 followed by 97 zeros). The approximate population of the earth is 5,900,000,000. (59 followed by 8 zeros). This results in 59,000,000,000 (59 followed by 9 zeros). The approximate chance of any two rolled fingerprints on earth being identical is $59/10^{88}$ or 59 chances in 1 followed by 88 zeros. With the exception of three consolidations, the most similar non-mate fingerprints are consistent with a random selection.

With the exception of three consolidations, the most similar non-mate fingerprints are consistent with a random selection.

EXPERIMENT 2 DESCRIPTION:

The second experiment compared a subset of the minutiae pattern of each of the fingerprints in the set with the minutia patterns of all of the fingerprints in the set. The subset of the minutiae pattern was extracted from the central 21.7% of the whole minutia pattern of the fingerprint and simulated a "latent" fingerprint extracted from a crime scene. The average area of 300 latent fingerprints is 21.7% of the average area of the 300 mates. This ratio was used in the extraction of the subset minutia patterns. The subset fingerprints are referred to as "search fingerprints" in the following discussion.

Experiment 2 required an additional two billion, five hundred million (2,500,000,000) comparisons.

The same software and analysis techniques were applied to the results of these comparisons with the following exception: the mate and most similar non-mate fingerprint Z scores were separated according to the number of minutiae in the search fingerprint (search minutia count), rather than being combined as in Experiment 1.

For each individual search minutia count we determined the probability that the degree of similarity of the mate with itself could occur by chance, given the set of 500 most similar fingerprints. Again, this was done using a statistical measure of separability known as "Sample Z Score". (The mathematical basis for this operation is described in "Statistics, Seventh Edition" - McClave, James T., et. al., Prentice Hall, ISBN 0-13-471542) The Z score can be converted into a probability using standard mathematics functions. (The result of these functions are tabulated in "Handbook of Mathematical Functions" - Abramowitz, Milton and Stegun, Irene A., Dover Publications. Library of Congress Catalog Card Number: 65-12253).

Significant Values of Z scores are:

Z	Probability	Explanation
0.0	1/2	Half the scores of a random sample will be higher (or lower).
3.0	1/500	Highest expected score in a random population of 500.
10.0	$1/10^{23}$	1 chance in 1 followed by 23 zeroes.
21.0	$1/10^{97}$	1 chance in 1 followed by 97 zeroes.

EXPERIMENT 2 RESULTS:

In the 50,000 simulated-latent print vs. 50,000 rolled print experiment, we discovered three fingerprints with unexpectedly high non-mate Z-scores. These were unlikely to occur and were a cause for concern:

Mate Index	Non-mate index	Z Score	Probability (approximate)
21583	21582	3.64	1/7,600
12640	21111	3.91	1/20,400
21111	12640	5.85	1/1,000,000,000

In one case, we found that the non-mate fingerprint had the same FBI Identification Number as did the mate:

21583	[redacted]T3
21582	[redacted]T3

This consolidation (high non-mate Z Score) was the result of different rollings of the same finger.

The other two cases the FBI Identification Numbers were different:

Index	FBI Identification Number
12640	[redacted]G
21111	[redacted]9

This called for further investigation.

We have supplied the FBI Identification Numbers to the Clarksburg, WV AFIS facility with the request that they recover the original 10-print card images and have a professional Fingerprint Examiner review them. The results of this review are not available as of 10:34, 27 March 1999.

We extracted the minutiae pattern data for all ten fingers with the subject FBI Identification Numbers from the Criminal Master File, for a total of 20 fingerprints. We used the same software as in the primary experiments to compare each of these prints with all of the others for a total of 400 comparisons. The results of this comparison are:

The highest scores for each comparison were (as expected) the subject fingerprint compared with itself. The second highest scores for each comparison corresponded to the finger position (in both directions):

[redacted]G	[redacted]9
Left Thumb	Left Thumb
Left Index	Left Index
Left Middle	Left Middle
Left Ring	Left Ring
Left Little	Left Little
Right Thumb	Right Thumb
Right Index	Right Index
Right Middle	Right Middle
Right Ring	Right Ring
Right Little	Right Little

The second highest scores were also significantly higher than the remaining scores.

This is highly indicative that the two FBI Identification Numbers refer to the same individual.

We removed these four fingerprints from the data set and produced the chart titled "Sample Z Score as a Function of Search Minutia Count". This chart shows that there is a trend in minutia pattern similarity such that the pattern similarity and the likelihood of an identical minutia pattern match occurring at random both decrease with increasing minutia count.

EXPERIMENT 2 CONCLUSIONS:

The probability of a minutia subset of a non-mate fingerprint being identical to a minutia subset of any particular fingerprint is less than $1/10^{27}$ (1 chance in 1 followed by 27 zeroes) for small numbers of minutiae (in this case, small means four), decreasing to less than $1/10^{97}$ (1 chance in 1 followed by 97 zeroes) for larger numbers of minutiae (in this case, larger means greater than eighteen).

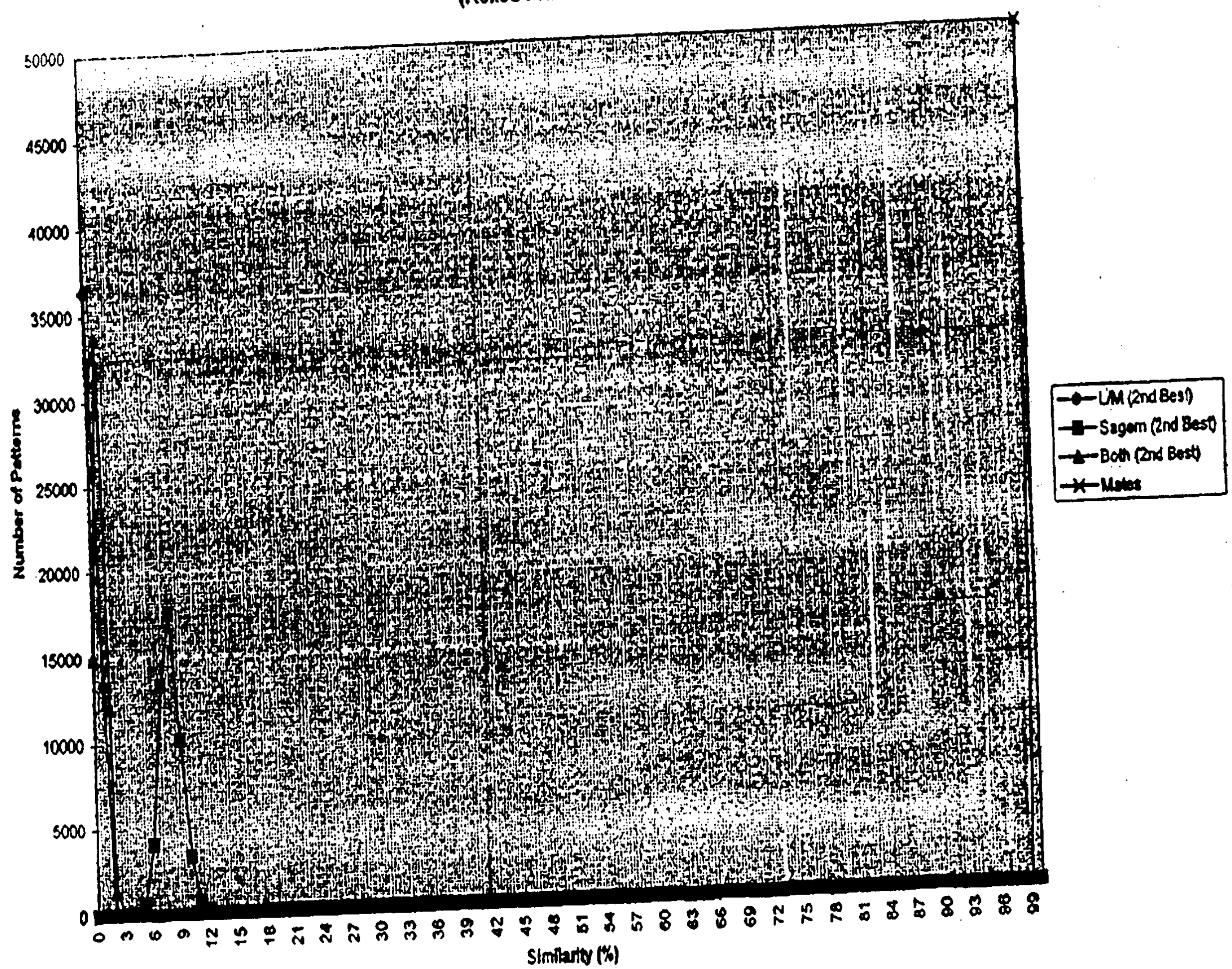
Using the same data as above for the population of the earth, the approximate chance of any two minutia subsets of fingerprints on earth being identical is $59/10^{16}$ or 59 chances in 1 followed by 18 zeros.

With thirteen (13) or more search minutiae, the most similar non-mate scores are consistent with a random selection.

With twelve (12) or fewer search minutiae, the most similar non-mate scores show that there is a chance of finding similar (but not identical) minutia patterns when comparing minutia subsets of non-mate fingerprints. This chance increases (as a general trend) with decreasing search minutia counts.

Prepared by AFIS LAB 4/5/99

Minutia Pattern Similarity
(Rolled Prints vs. Rolled Prints)



core

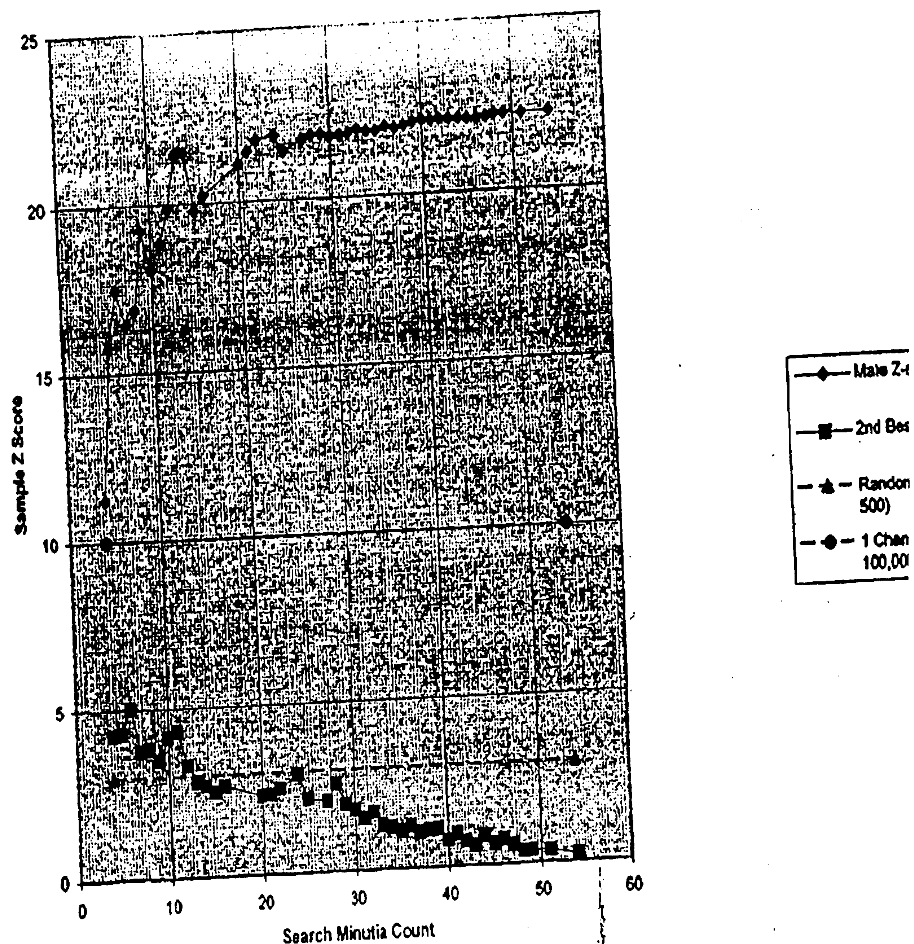
Z-score

Chance (1 in 500 out of

in
000,000,000,000,000,000

Prepared by AFIS LAB 4/5/99

Sample Z Score as a Function of Search Minutia Count



Karen Kafadar, "Statistical Issues in Assessing Forensic Evidence," (2015) International Statistical Review (2015), 83, 1, 111, online at: http://onlinelibrary.wiley.com/doi/10.1111/insr.12069/abstract

Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts, "Accuracy and Reliability of Forensic Latent Fingerprint Decisions," (2011) 108 PNAS 7733-7738.

Accuracy and reliability of forensic latent fingerprint decisions

Bradford T. Ulery[a], R. Austin Hicklin[a], JoAnn Buscaglia[b,1], and Maria Antonia Roberts[c]

[a]Noblis, 3150 Fairview Park Drive, Falls Church, VA 22042; [b]Counterterrorism and Forensic Science Research Unit, Federal Bureau of Investigation Laboratory Division, 2501 Investigation Parkway, Quantico, VA 22135; and [c]Latent Print Support Unit, Federal Bureau of Investigation Laboratory Division, 2501 Investigation Parkway, Quantico, VA 22135

Edited by Stephen E. Fienberg, Carnegie Mellon University, Pittsburgh, PA, and approved March 31, 2011 (received for review December 16, 2010)

The interpretation of forensic fingerprint evidence relies on the expertise of latent print examiners. The National Research Council of the National Academies and the legal and forensic sciences communities have called for research to measure the accuracy and reliability of latent print examiners' decisions, a challenging and complex problem in need of systematic analysis. Our research is focused on the development of empirical approaches to studying this problem. Here, we report on the first large-scale study of the accuracy and reliability of latent print examiners' decisions, in which 169 latent print examiners each compared approximately 100 pairs of latent and exemplar fingerprints from a pool of 744 pairs. The fingerprints were selected to include a range of attributes and quality encountered in forensic casework, and to be comparable to searches of an automated fingerprint identification system containing more than 58 million subjects. This study evaluated examiners on key decision points in the fingerprint examination process; procedures used operationally include additional safeguards designed to minimize errors. Five examiners made false positive errors for an overall false positive rate of 0.1%. Eighty-five percent of examiners made at least one false negative error for an overall false negative rate of 7.5%. Independent examination of the same comparisons by different participants (analogous to blind verification) was found to detect all false positive errors and the majority of false negative errors in this study. Examiners frequently differed on whether fingerprints were suitable for reaching a conclusion.

The interpretation of forensic fingerprint evidence relies on the expertise of latent print examiners. The accuracy of decisions made by latent print examiners has not been ascertained in a large-scale study, despite over one hundred years of the forensic use of fingerprints. Previous studies (1–4) are surveyed in ref. 5. Recently, there has been increased scrutiny of the discipline resulting from publicized errors (6) and a series of court admissibility challenges to the scientific basis of fingerprint evidence (e.g., 7–9). In response to the misidentification of a latent print in the 2004 Madrid bombing (10), a Federal Bureau of Investigation (FBI) Laboratory review committee evaluated the scientific basis of friction ridge examination. That committee recommended research, including the study described in this report: a test of the performance of latent print examiners (11). The need for evaluations of the accuracy of fingerprint examination decisions has also been underscored in critiques of the forensic sciences by the National Research Council (NRC, ref. 12) and others (e.g., refs. 13–16).

Background

Latent prints ("latents") are friction ridge impressions (fingerprints, palmprints, or footprints) left unintentionally on items such as those found at crime scenes (*SI Appendix*, *Glossary*). Exemplar prints ("exemplars"), generally of higher quality, are collected under controlled conditions from a known subject using ink on paper or digitally with a livescan device (17). Latent print examiners compare latents to exemplars, using their expertise rather than a quantitative standard to determine if the information content is sufficient to make a decision. Latent print examination can be complex because latents are often small, unclear, distorted, smudged, or contain few features; can overlap with other prints or appear on complex backgrounds; and can contain artifacts from the collection process. Because of this complexity, experts must be trained in working with the various difficult attributes of latents.

During examination, a latent is compared against one or more exemplars. These are generally collected from persons of interest in a particular case, persons with legitimate access to a crime scene, or obtained by searching the latent against an Automated Fingerprint Identification System (AFIS), which is designed to select from a large database those exemplars that are most similar to the latent being searched. For latent searches, an AFIS only provides a list of candidate exemplars; comparison decisions must be made by a latent print examiner. Exemplars selected by an AFIS are far more likely to be similar to the latent than exemplars selected by other means, potentially increasing the risk of examiner error (18).

The prevailing method for latent print examination is known as analysis, comparison, evaluation, and verification (ACE-V) (19, 20). The ACE portion of the process results in one of four decisions: the analysis decision of no value (unsuitable for comparison); or the comparison/evaluation decisions of individualization (from the same source), exclusion (from different sources), or inconclusive. The Scientific Working Group on Friction Ridge Analysis, Study and Technology guidelines for operational procedures (21) require verification for individualization decisions, but verification is optional for exclusion or inconclusive decisions. Verification may be blind to the initial examiner's decision, in which case all types of decisions would need to be verified. ACE-V has come under criticism by some as being a general approach that is underspecified (e.g., refs. 14 and 15).

Latent-exemplar image pairs collected under controlled conditions for research are known to be mated (from the same source) or nonmated (from different sources). An individualization decision based on mated prints is a true positive, but if based on nonmated prints, it is a false positive (error); an exclusion decision based on mated prints is a false negative (error), but is a true negative if based on nonmated prints. The term "error" is used in this paper only in reference to false positive and false negative conclusions when they contradict known ground truth. No such absolute criteria exist for judging whether the evidence is sufficient to reach a conclusion as opposed to making an inconclusive or no-value decision. The best information we have to

APPLIED BIOLOGICAL SCIENCES

Author contributions: B.T.U., R.A.H., J.B., and M.A.R. designed research; B.T.U., R.A.H., J.B., and M.A.R. performed research; B.T.U. and R.A.H. contributed new analytic tools; B.T.U., R.A.H., J.B., and M.A.R. analyzed data; and B.T.U., R.A.H., J.B., and M.A.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: joann.buscaglia@ic.fbi.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1018707108/-/DCSupplemental.

evaluate the appropriateness of reaching a conclusion is the collective judgments of the experts. Various approaches have been proposed to define sufficiency in terms of objective minimum criteria (e.g., ref. 22), and research is ongoing in this area (e.g., ref. 23). Our study is based on a black box approach, evaluating the examiners' accuracy and consensus in making decisions rather than attempting to determine or dictate how those decisions are made (11, 24).

Study Description

This study is part of a larger research effort to understand the accuracy of examiner conclusions, the level of consensus among examiners on decisions, and how the quantity and quality of image features relate to these outcomes. Key objectives of this study were to determine the frequency of false positive and false negative errors, the extent of consensus among examiners, and factors contributing to variability in results. We designed the study to enable additional exploratory analyses and gain insight in support of the larger research effort.

There is substantial variability in the attributes of latent prints, in the capabilities of latent print examiners, in the types of casework received by agencies, and the procedures used among agencies. Average measures of performance across this heterogeneous population are of limited value (25)—but do provide insight necessary to understand the problem and scope future work. Furthermore, there are currently no means by which all latent print examiners in the United States could be enumerated or used as the basis for sampling: A representative sample of latent print examiners or casework is impracticable.

To reduce the problem of heterogeneity, we limited our scope to a study of performance under a single, operationally common scenario that would yield relevant results. This study evaluated examiners at the key decision points during analysis and evaluation. Operational latent print examination processes may include additional steps, such as examination of original evidence or paper fingerprint cards, review of multiple exemplars from a subject, consultation with other examiners, revisiting difficult comparisons, verification by another examiner, and quality assurance review. These steps are implemented to reduce the possibility of error.

Ideally, a study would be conducted in which participants were not aware that they were being tested. The practicality of such an approach even within a single organization would depend on the type of casework. Fully electronic casework could allow insertion of test data into actual casework, but this may be complex to the point of infeasibility for agencies in which most examinations involve physical evidence, especially when chain-of-custody issues are considered. Combining results among multiple agencies with heterogeneous procedures and types of casework would be problematic.

In order to get a broad cross-section of the latent print examiner community, participation was open to practicing latent print examiners from across the fingerprint community. A total of 169 latent print examiners participated; most were volunteers, while the others were encouraged or required to participate by their employers. Participants were diverse with respect to organization, training history, and other factors. The latent print examiners were generally highly experienced: Median experience was 10 y, and 83% were certified as latent print examiners. More detailed descriptions of participants, fingerprint data, and study procedures are included in *SI Appendix, Materials and Methods*.

The fingerprint data included 356 latents, from 165 distinct fingers from 21 people, and 484 exemplars. These were combined to form 744 distinct latent-exemplar image pairs. There were 520 mated and 224 nonmated pairs. The number of fingerprint pairs used in the study, and the number of examiners assigned to each pair, were selected as a balance between competing research priorities: Measuring consensus and variability among examiners required multiple examiners for each image pair, while incorporating a broad range of fingerprints for measuring image-specific effects required a large number of images.

We sought diversity in fingerprint data, within a range typical of casework. Subject matter experts selected the latents and mated exemplars from a much larger pool of images to include a broad range of attributes and quality. Latents of low quality were included in the study to evaluate the consensus among examiners in making value decisions about difficult latents. The exemplar data included a larger proportion of poor-quality exemplars than would be representative of exemplars from the FBI's Integrated AFIS (IAFIS) (*SI Appendix*, Table S4). Image pairs were selected to be challenging: Mated pairs were randomly selected from the multiple latents and exemplars available for each finger position; nonmated pairs were based on difficult comparisons resulting from searches of IAFIS, which includes exemplars from over 58 million persons with criminal records, or 580 million distinct fingers (*SI Appendix*, section 1.3). Participants were surveyed, and a large majority of the respondents agreed that the data were representative of casework (*SI Appendix*, Table S3).

Noblis developed custom software for this study in consultation with latent print examiners, who also assessed the software and test procedures in a pilot study. The software presented latent and exemplar images to the participants, allowed a limited amount of image processing, and recorded their decisions, as indicated in Fig. 1 (*SI Appendix*, section 1.2). Each of the examiners was randomly assigned approximately 100 image pairs out of the total pool of 744 image pairs (*SI Appendix*, section 1.3). The image pairs were presented in a preassigned order; examiners could not revisit previous comparisons. They were given several weeks to complete the test. Examiners were instructed to use the same diligence that they would use in performing casework. Participants were assured that their results would remain anonymous; a coding system was used to ensure anonymity during analysis and in reporting.

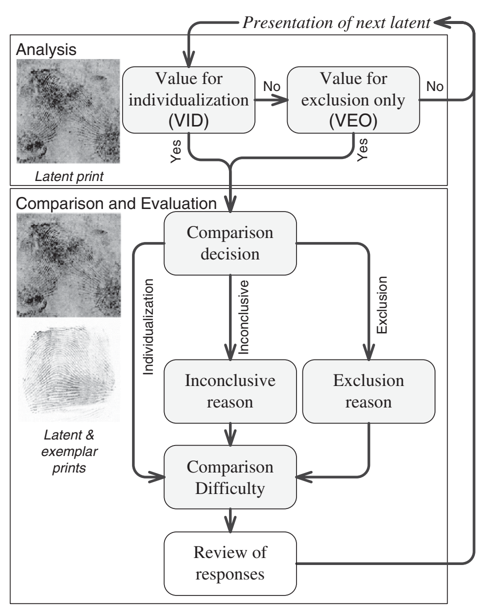


Fig. 1. Software workflow. Each examiner was assigned a distinct, randomized sequence of image pairs. For each pair, the latent was presented first for a value decision; if it was determined to be no value, the test proceeded directly to the latent from the next image pair; otherwise, an exemplar was presented for comparison and evaluation (*SI Appendix*, section 1.5).

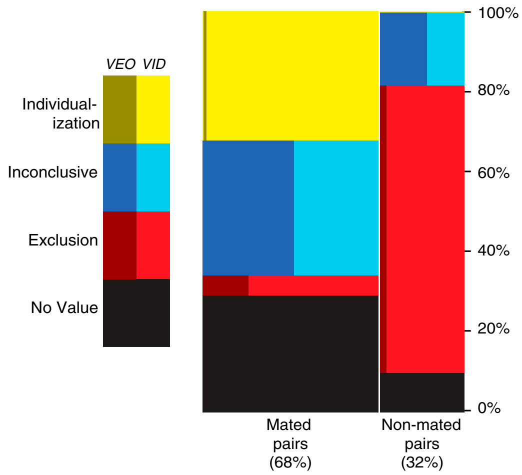


Fig. 2. Distribution of 17,121 decisions. 23% of all decisions resulted in no-value decisions (no comparison was performed); comparison decisions were based on latents of VID and of VEO; 7.5% of comparisons of mated pairs resulted in exclusion decisions (false negatives); 0.1% of comparisons of nonmated pairs resulted in individualization decisions (false positives—too few to be visible) (*SI Appendix*, Table S5).

Results

A summary of examiner decisions is shown in Fig. 2. We emphasize that individual examiner decisions are only a part of an overall operational process, which may include verification, quality assurance, and reporting. Our results do not necessarily reflect the performance of this overall operational process.

The true negative rate was greater than the true positive rate. Much of this difference may be explained by three factors: The amount of information necessary for an exclusion decision is typically less than for an individualization decision, examiners operate within a culture where false positives are seen as more serious errors than false negatives (5), and the mated pairs included a greater proportion of poor-quality prints than the nonmated pairs (*SI Appendix*, section 1.3). Whereas poor-quality latents result in the no-value decisions in Fig. 2, the poor-quality exemplars contribute to an increase in the proportion of inconclusive decisions.

Rates of comparison decisions can be calculated as a percentage of all presentations (PRES), including latents of no value; of comparisons where the latent was of value for individualization (VID); or of all comparisons (CMP), which includes comparisons where the latent was of value for exclusion only (VEO) as well as VID. Because standard operating procedures typically include only VID comparisons, this is our default basis for reporting these rates.

False Positives

Six false positives occurred among 4,083 VID comparisons of nonmated pairs (false positive rate, $FPR_{VID} = 0.1\%$) (*SI Appendix*, Tables S5 and S8; confidence intervals are discussed in *SI Appendix*, section 2.1). The image pairs that resulted in two of the false positives are shown in Fig. 3. Two of the false positive errors involved a single latent, but with exemplars from different subjects. Four of the five distinct latents on which false positives occurred (vs. 18% of nonmated latents) were deposited on a galvanized metal substrate, which was processed with cyanoacrylate and light gray powder. These images were often partially or fully tonally reversed (light ridges instead of dark), on a complex background (Fig. 3, image pair C). It is not known if other complex backgrounds or processing artifacts would have a similar increased potential for error.

The six errors were committed by five examiners, three of whom were certified (including one examiner who made two errors); one was not certified; one did not respond to our background survey. These correspond to the overall proportions of certifications among participants (*SI Appendix*, section 1.4). In no case did two examiners make the same false positive error: Five errors occurred on image pairs where a large majority of examiners correctly excluded; one occurred on a pair where the majority of examiners made inconclusive decisions. This suggests that these erroneous individualizations would have been detected if blind verification were routinely performed. For verification to be truly blind, examiners must not know that they are verifying individualizations; this can be ensured by performing verifications on a mix of conclusion types, not merely individualizations. The general consensus among examiners did not indicate that these were difficult comparisons, and only for two of the six false positives did the examiner making the error indicate that these were difficult (*SI Appendix*, Table S8).

There has been discussion (24, 26, 27) regarding the appropriateness of using qualified conclusions in investigation or testimony. The effects of qualified conclusions could be assessed in this study, as "inconclusive with corresponding features" (*SI Appendix*, section 1.5). Qualified conclusions potentially yield many additional "leads": 36.5% of VID comparisons resulted in individualization decisions, and an additional 6.2% resulted in qualified conclusions. However, 99.8% of individualization decisions were mated, as opposed to only 80.6% of qualified conclusions (*SI Appendix*, section 2). Only one of the six image pairs

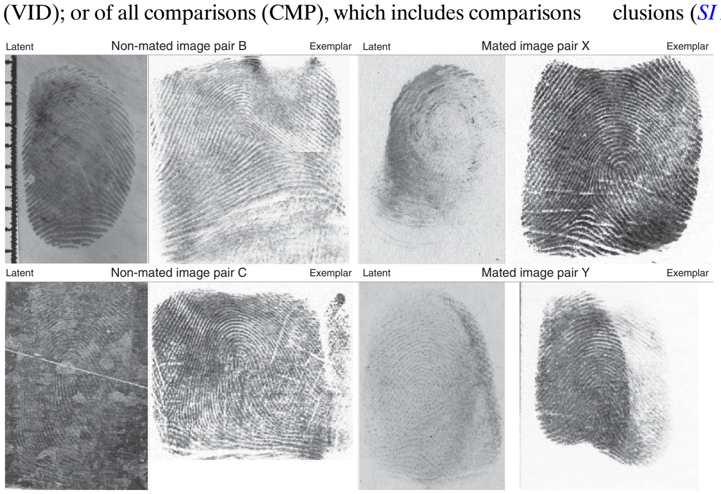


Fig. 3. Examples of fingerprint pairs used in the study that resulted in examiner errors. Pairs B and C resulted in false positive errors: 1 of 30 examiners made an individualization decision on B (24 exclusions); 1 of 26 examiners made an individualization decision on C (22 exclusions). The processing of the latent in C (cyanoacrylate with light gray powder) tonally reversed the image so that portions of ridges were light rather than dark. Pairs X and Y resulted in false negative errors, with no true positives made by any examiner: X was excluded by 13 of 29 examiners, presumably because the latent was deposited with a twisting motion that resulted in misleading ridge flow; Y was excluded by 15 of 18 examiners; the exemplar was particularly distorted. For use in this figure, these images were cropped to reduce background area.

that resulted in false positives had a plurality of inconclusive decisions, and none had a plurality "with corresponding features."

False Negatives

False negatives were much more prevalent than false positives (false negative rate: $FNR_{VID} = 7.5\%$) (*SI Appendix*, Table S5). Including VEO comparisons had no substantial effect: $FNR_{CMP} = 7.5\%$. Eighty-five percent of examiners made at least one false negative error, despite the fact that 65% of participants said that they were unaware of ever having made an erroneous exclusion after training (*SI Appendix*, section 1.4, no. 25); awareness of previous errors was not correlated with false negative errors on this test. False negatives were distributed across half of the image pairs that were compared. The likelihood of false negatives varied significantly by examiner (discussed further under *Examiner Skill*, below), and by image pair (*SI Appendix*, Figs. S3 and S5 *C* and *D*). Of the image pairs that were most frequently associated with false negatives, most had distorted latents and/or exemplars that gave an appearance of a different ridge flow pattern.

Verification of exclusions (especially blind verification) is not standard practice in many organizations, in part due to the large number encountered in casework. To investigate the potential benefits of blind verification, we posed the following question: Given a mated image pair, what is the probability, p_v, that two examiners would both reach exclusion decisions? If exclusions were equally likely for all image pairs (independence assumption), we would estimate that exclusions by two examiners would occur at the rate $p_v = FNR^2_{PRES} = 5.3\% \times 5.3\% = 0.3\%$ (*SI Appendix*, Table S5). However, the data show that the independence assumption is not valid: Some mated pairs are more likely to be excluded than others. Because the outcomes of blind verifications are not statistically independent but depend on the image pairs, we estimate $p_v = 0.85\%$ (*SI Appendix*, section 11). This suggests that blind verification of exclusions could greatly reduce false negative errors; agency policy would have to balance this benefit with the impact on limited resources.

For exclusions where the latent was VID, examiner assessment of comparison difficulty was a good predictor of accuracy, but even "Very Easy/Obvious" exclusions were sometimes incorrect: Among 450 false negatives where the latent was VID, 13 were rated "Very Easy/Obvious" by 11 distinct examiners (*SI Appendix*, Fig. S8). Latent value (VEO vs. VID) had no predictive value for false negative errors; however, exclusions were more likely to be true negatives when the latent was VID than when it was VEO. This counterintuitive result is due to the fact that VEO determinations were more often inconclusive, hence most exclusion decisions were associated with VID latents (*SI Appendix*, Fig. S7).

Posterior Probabilities

False positive and false negative rates are important accuracy measures, but assume a priori knowledge of true mating relationships, which of course are not known in forensic casework. In practice, knowledge of mating relationships is based solely on examiners' decisions: It is important to know the likelihood that these decisions are correct. Positive predictive value (PPV) is the percentage of individualization decisions that are true positives; negative predictive value (NPV) is the percentage of exclusion decisions that are true negatives. Fig. 4 depicts PPV and NPV as functions of the prior prevalence of mated pairs among the examinations performed: As the proportion of mated pairs increases, PPV increases and NPV decreases (*SI Appendix*, section 9). The prior prevalence of mated pair comparisons varies substantially among organizations, by case type, and by how candidates are selected. Mated comparisons are far more prevalent in cases where the candidates are suspects determined by nonfingerprint means than in cases where candidates were selected by an AFIS.

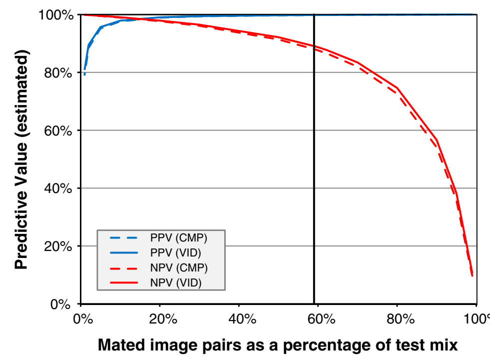


Fig. 4. PPV and NPV as a function of mate prevalence in workload. The observed predictive values ($PPV_{VID,59\%} = 99.8\%$ and $NPV_{VID,59\%} = 88.9\%$ for VID comparisons) correspond to the actual test mix (indicated) where 59% of VID comparisons were mated pairs; other predictive values are calculated as a function of mate prevalence. Sixty-two percent of all comparisons (VEO and VID) were performed on mated pairs, and $PPV_{CMP,62\%} = 99.8\%$ and $NPV_{CMP,62\%} = 86.6\%$.

Consensus

Each image pair was examined by an average of 23 participants. Their decisions can be regarded as votes in a decision space (Fig. 5). Consensus was limited on both mated and nonmated pairs: VID decisions were unanimous on 48% of mated pairs and 33% of nonmated pairs. Votes by latent print examiners also provide a basis for assessing sufficiency for value decisions, as shown in Fig. 6; consensus on individualization and exclusion decisions is shown in *SI Appendix*, Fig. S6.

Lack of consensus among examiners can be attributed to several factors. For unanimous decisions, the images were clearly the driving factor: Unusable or pristine prints resulted in unanimous decisions, and therefore different data selection would have affected the extent of consensus. When there was a lack of consensus, much of the variation could be explained by examiner differences: Examiners showed varying tendencies toward no-

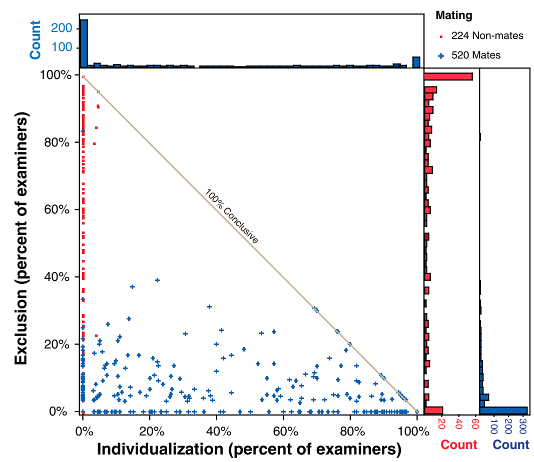


Fig. 5. Decision rates on each image pair. Percentage of examiners making an individualization decision (*x* axis) vs. exclusion decision (*y* axis) on each image pair; mean 23 presentations per pair. VEO and no-value decisions are treated as inconclusive. Marginal distributions are shown as histograms. Of mated pair decisions, 10% were unanimous true positives, 38% unanimous inconclusives. Of nonmated pair decisions, 25% were unanimous true negatives, 9% were unanimous inconclusives. Points along diagonal represent pairs on which all examiners reached conclusions. The prevalence of false negatives is evident in the vertical spread of mated pairs; the few false positives are evident in the limited horizontal spread of the nonmated pairs.

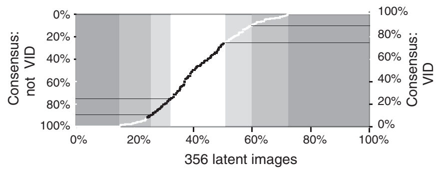


Fig. 6. Examiner consensus on VID decisions, showing the percentage of examiners reaching consensus (*y* axis) on each latent (*x* axis). Areas of unanimous (100%), decile (10%, 90%), and quartile (25%, 75%) consensus are marked. For example, at a 90% level of consensus (*y* axes), examiners agreed that 40% of the latents were VID (interval from 60% to 100% indicated by a horizontal line in upper right) (*SI Appendix*, Table S11). Such measures of consensus may be useful in developing quantity and quality metrics.

value or inconclusive decisions, or toward conclusions (*SI Appendix*, Fig. S4). Examiners differed significantly in conclusion rates, and we see this effect as secondary to image characteristics in explaining lack of consensus. Other factors accounting for lack of consensus include intraexaminer inconsistency and (presumably) test environment (*SI Appendix*, Fig. S3).

It was not unusual for one examiner to render an inconclusive decision while another made an individualization decision on the same comparison. This result is consistent with previous observations (1, 5, 28). Among all decisions based on mated pairs, 23.0% resulted in decisions other than individualization even though at least one other examiner made a true positive on the same image pair; 4.8% were not individualization decisions even though the majority of other examiners made true positives. This has operational implications in that some potential individualizations are not being made, and contradictory decisions are to be expected.

When examiners reached contradictory conclusions (exclusion and individualization) on a single comparison, the exclusion decision was more frequently in error: 7.7% of independent examinations of conclusions on mates were contradictory, vs. 0.23% on nonmates. Which of the contradictory decisions is more likely to be erroneous depends on the prior prevalence of mated vs. nonmated pairs: Exclusion decisions are more likely to be erroneous except in situations where the prior prevalence of nonmated pairs is very high.

Examiner Skill

The criminal justice system relies on the skill of latent print examiners as expert witnesses. Currently, there is no generally accepted objective measure to assess the skill of latent print examiners. Skill is multidimensional and is not limited to error rates (FPR and FNR), but also includes TPR, true negative rate (TNR), VID and VEO rates, and conclusion rate (CR—the percentage of individualization or exclusion conclusions as opposed to no-value or inconclusive decisions). Any assessment of skill must consider these dimensions. Although most discussions of examiner skill focus on error rates (e.g., ref. 13), the other aspects of examiner skill are important not just to the examiner's organization, but to the criminal justice system as well; e.g., an examiner who is frequently inconclusive is ineffective and thereby fails to serve justice. Both individual examiners and organizations must strike a proper balance between the societal costs of errors and inappropriate decisions, and the operational costs of detection. Contradictory verification decisions, whether involving erroneous conclusions or inappropriate inconclusive decisions, should be internally documented and addressed through an organization's continual improvement processes.

We found that examiners differed substantially along these dimensions of skill, and that these dimensions were largely independent. Our study measured all of these dimensions with the exception of FPRs for individual examiners, which were too low to measure with precision (*SI Appendix*, section 3). Fig. 7 shows that examiners' conclusion rates (CR_{PRES}) varied from 15 to 64% (mean 37%, SD 10%) on mated pairs, and from 7 to 96% (mean

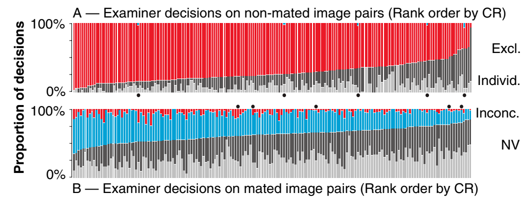


Fig. 7. Decision rates by examiner. Proportions of decisions for all 169 examiners on (*A*) nonmated and (*B*) mated image pairs. Examiners in each chart are sorted on CR. Each examiner was randomly assigned 51 to 74 mated image pairs (mean 69, SD 5) and 26 to 53 nonmated image pairs (mean 33, SD 7). In both, errors are shown in red. Column width indicates the number of image pairs. Examiners who made false positive errors are indicated with black dots (*SI Appendix*, Table S7).

71%, SD 14%) on nonmated pairs. The observed range in CRs may be explained by a higher level of skill (ability to reach more conclusions at the same level of accuracy), or it may imply a higher risk tolerance (more conclusions reached at the expense of making more errors).

Fig. 7 shows substantial variability in CR among examiners. These measured rates were based on an average of 69 mated presentations and 33 nonmated presentations. The limited number of presentations resulted in a wide margin of measurement error when evaluating the performance of an individual examiner (*SI Appendix*, Fig. S5). Although the estimates for each examiner are statistically unbiased, the sampling error in these estimates contributed substantially to the observed variability among examiners. The observed variability is a biased estimate that overstates the true variability (*SI Appendix*, Figs. S3*B* and S4).

Fig. 8 shows the relations between three of the skill dimensions measured for each examiner. Blue squares near the lower right of the chart represent highly skilled examiners: accurate (making few or no errors) and effective (high TNR and TPR, and therefore high CR). The red cross at the bottom left denotes an accurate (0% FNR_{VID}), but ineffective (5% TNR_{VID}, 16% TPR_{PRES}) examiner. The examiner denoted by the red cross at the top right is inaccurate (34% FNR_{VID}), and has mixed effectiveness (100% TNR_{VID}, 23% TPR_{PRES}). Attempting to compare the skill of any two examiners is a multidimensional problem. A combination of multiple dimensions into a single hypothetical measure of skill would require a weighting function to trade off the relative value of each dimension; such weighting might be driven by policy, based on the relative cost/benefit of each dimension for operational needs.

Tests could be designed to measure examiner skill along the multiple dimensions discussed here. Such tests could be valuable

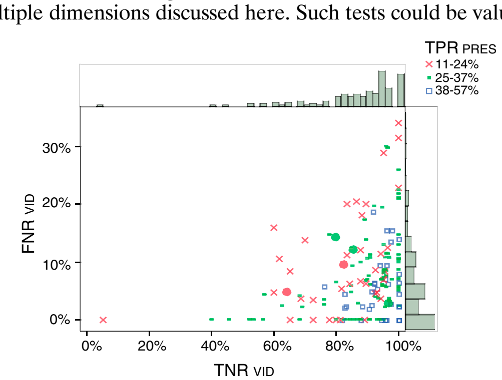


Fig. 8. Examiner skill. Each of the 169 examiners is plotted on three skill dimensions: TNR_{VID} (mean 88%, SD 13.6%), FNR_{VID} (mean 7.5%, SD 7.3%), and TPR_{PRES} (shown in color, with red crosses denoting the lowest quartile and blue squares the highest quartile; mean 32%, SD 9.4%). The five examiners who made false positive errors are indicated with bold filled circles.

not just as traditional proficiency tests with pass/fail thresholds, but as a means for examiners or their organizations to understand skills for specific training, or for tasking based on skills (such as selecting examiners for verification based on complementary skill sets).

Certified examiners had higher conclusion rates than noncertified examiners without a significant change in accuracy (significantly higher TPR_{VID} and TNR_{VID}; FNR_{VID} did not vary significantly) (*SI Appendix*, section 6). Length of experience as a latent print examiner did not show a significant correlation with TPR_{VID}, TNR_{VID}, or FNR_{VID} (*SI Appendix*, Table S9 and Fig. S2).

Examiners with a lower TPR_{VID} tended also to have a lower TNR_{VID}. Examiners with a higher FNR_{VID} tended to have a lower TPR_{VID}. Examiners with a higher TNR_{VID} tended also to have a higher FNR_{VID} (*SI Appendix*, Table S9 and Fig. S2).

Conclusions

Assessing the accuracy and reliability of latent print examiners is of great concern to the legal and forensic science communities. We evaluated the accuracy of decisions made by latent print examiners on difficult fingerprint comparisons in a computer-based test corresponding to one stage in AFIS casework. The rates measured in this study provide useful reference estimates that can inform decision making and guide future research; the results are not representative of all situations, and do not account for operational context and safeguards. False positive errors (erroneous individualizations) were made at the rate of 0.1% and never by two examiners on the same comparison. Five of the six errors occurred on image pairs where a large majority of examiners made true negatives. These results indicate that blind verification should be highly effective at detecting this type of error. Five of the 169 examiners (3%) committed false positive errors, out of an average of 33 nonmated pairs per examiner.

False negative errors (erroneous exclusions) were much more frequent (7.5% of mated comparisons). The majority of examiners (85%) committed at least one false negative error, with individual examiner error rates varying substantially, out of an average of 69 mated pairs per examiner. Blind verification would have detected the majority of the false negative errors; however, verification of exclusion decisions is not generally practiced in operational procedures, and blind verification is even less frequent. Policymakers will need to consider tradeoffs between the financial and societal costs and benefits of additional verifications.

Most of the false positive errors involved latents on the most complex combination of processing and substrate included in the study. The likelihood of false negatives also varied by image. Further research is necessary to identify the attributes of prints associated with false positive or false negative errors, such as quality, quantity of features, distortion, background, substrate, and processing method.

Examiners reached varied levels of consensus on value and comparison decisions. Although there is currently no objective basis for determining the sufficiency of information necessary to reach a fingerprint examination decision, further analysis of the data from this study will assist in defining quality and quantity metrics for sufficiency. This lack of consensus for comparison decisions has a potential impact on verification: Two examiners will sometimes reach different conclusions on a comparison.

Examiner skill is multidimensional and is not limited to error rates. Examiner skill varied substantially. We measured various dimensions of skill and found them to be largely independent.

This study is part of a larger ongoing research effort. To further our understanding of the accuracy and reliability of latent print examiner decisions, we are developing fingerprint quality and quantity metrics and analyzing their relationship to value and comparison decisions; extending our analyses to include detailed examiner markup of feature correspondence; collecting fingerprints specifically to explore how complexity of background, substrate and processing are related to comparison decisions; and measuring intraexaminer repeatability over time.

This study addresses in part NRC Recommendation 3 (12), developing and quantifying measures of accuracy and reliability for forensic analyses, and will assist in supporting the scientific basis of forensic fingerprint examination. The results of this study will provide insight into developing operational procedures and training of latent print examiners and will aid in the experimental design of future proficiency tests of latent print examiners.

ACKNOWLEDGMENTS. We thank the latent print examiners who participated in this study, as well as William Fellner, Jill McCracken, Keith Ward, Stephen Meagher, Calvin Yeung, Ted Unnikumaran, Erik Stanford, and William Chapman. This is publication number 10-19 of the FBI Laboratory Division. This work was funded in part under a contract award to Noblis, Inc. from the FBI Biometric Center of Excellence and in part by the FBI Laboratory Division. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the US government.

1. Evett IW, Williams RL (1995) A review of the 16 point fingerprint standard in England and Wales. *Fingerprint Whorld* 21.
2. Wertheim K, Langenburg G, Moenssens A (2006) A report of latent print examiner accuracy during comparison training exercises. *J Forensic Identification* 56:55–93.
3. Gutowski S (2006) Error rates in fingerprint examination: The view in 2006. *Forensic Bulletin* 2006:18–19 Autumn.
4. Langenburg G, Champod P, Wertheim P (2009) Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons. *J Forensic Sci* 54:571–582.
5. Langenburg G (2009) A performance study of the ACE-V process. *J Forensic Identification* 59:219–257.
6. Cole SA (2005) More than zero: Accounting for error in latent fingerprint identification. *J Crim Law Criminol* 95:985–1078.
7. *United States v Mitchell* No. 96-407 (ED PA 1999).
8. *United States v Llera Plaza* Cr. No. 98-362-10, 11, 12 (ED PA 2002).
9. *Maryland v Rose* No. K06-0545 (MD Cir 2007).
10. Office of the Inspector General (2006) *A Review of the FBI's Handling of the Brandon Mayfield Case* (US Department of Justice, Washington, DC).
11. Budowle B, Buscaglia J, Perlman RS (2006) Review of the scientific basis for friction ridge comparisons as a means of identification: Committee findings and recommendations. *Forensic Sci Commun* 8:1.
12. National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, Washington, DC).
13. Koehler JJ (2008) Fingerprint error rates and proficiency tests: What they are and why they matter. *Hastings Law J* 59:1077–1110.
14. Mnookin JL (2008) The validity of latent fingerprint identification: Confessions of a fingerprinting moderate. *Law Probability and Risk* 7:127–141.
15. Haber L, Haber RN (2008) Scientific validation of fingerprint evidence under Daubert. *Law Probability and Risk* 7:87–109.
16. Cole S (2006) Is fingerprint identification valid? Rhetorics of reliability in fingerprint proponents' discourse. *Law Policy* 28:109–135.
17. Scientific Working Group on Friction Ridge Analysis, Study and Technology (2011) Standard terminology of friction ridge examination, Version 3., Available at http://www.swgfast.org/documents/terminology/110323_Standard-Terminology_3.0.pdf.
18. Dror I, Mnookin J (2010) The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law Probability and Risk* 9:47–67.
19. Huber RA (1959) Expert witness. *Criminal Law Quarterly* 2:276–296.
20. Ashbaugh D (1999) *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology* (CRC Press, New York).
21. Scientific Working Group on Friction Ridge Analysis, Study and Technology (2002) Friction ridge examination methodology for latent print examiners, Version 1.01., Available at http://www.swgfast.org/documents/methodology/100506-Methodology-Reformatted-1.01.pdf.
22. Champod C (1995) Edmond Locard—Numerical standards and "probable" identifications. *J Forensic Identification* 45:136–155.
23. Neumann C, et al. (2007) Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *J Forensic Sci* 52:54–64.
24. Mnookin JL (2008) Of black boxes, instruments, and experts: Testing the validity of forensic science. *Episteme* 5:343–358.
25. Budowle B, et al. (2009) A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. *J Forensic Sci* 54:798–809.
26. Saks M, Koehler J (2005) The coming paradigm shift in forensic identification science. *Science* 309:892–895.
27. Stoney DA (1991) What made us ever think we could individualize using statistics? *J Forensic Sci Soc* 31:197–199.
28. Grieve DL (1996) Possession of truth. *J Forensic Identification* 46:521–528.

CLASS 2

Federal Rules of Evidence, 28 U.S.C.A., Rule 702. Testimony by Expert Witness. Document in public domain.

United States Code Annotated
Federal Rules of Evidence Rule 702, 28 U.S.C.A.
Rule 702. Testimony by Expert Witnesses

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

(a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;

(b) the testimony is based on sufficient facts or data;

(c) the testimony is the product of reliable principles and methods; and

(d) the expert has reliably applied the principles and methods to the facts of the case.

NOTES

(Pub. L. 93–595, §1, Jan. 2, 1975, 88 Stat. 1937; Apr. 17, 2000, eff. Dec. 1, 2000; Apr. 26, 2011, eff. Dec. 1, 2011.)

NOTES OF ADVISORY COMMITTEE ON PROPOSED RULES

An intelligent evaluation of facts is often difficult or impossible without the application of some scientific, technical, or other specialized knowledge. The most common source of this knowledge is the expert witness, although there are other techniques for supplying it.

Most of the literature assumes that experts testify only in the form of opinions. The assumption is logically unfounded. The rule accordingly recognizes that an expert on the stand may give a dissertation or exposition of scientific or other principles relevant to the case, leaving the trier of fact to apply them to the facts. Since much of the criticism of expert testimony has centered upon the hypothetical question, it seems wise to recognize that opinions are not indispensable and to encourage the use of expert testimony in non-opinion form when counsel believes the trier can itself draw the requisite inference. The use of opinions is not abolished by the rule, however. It will continue to be permissible for the experts to take the further step of suggesting the inference which should be drawn from applying the specialized knowledge to the facts. See Rules 703 to 705.

Whether the situation is a proper one for the use of expert testimony is to be determined on the basis of assisting the trier. "There is no more certain test for determining when experts may be used than the common sense inquiry whether the untrained layman would be qualified to determine intelligently and to the best possible degree the particular issue without enlightenment from those having a specialized understanding of the subject involved in the dispute." Ladd, Expert Testimony, 5 Vand.L.Rev. 414, 418 (1952). When opinions are excluded, it is because they are unhelpful and therefore superfluous and a waste of time. 7 Wigmore §1918.

The rule is broadly phrased. The fields of knowledge which may be drawn upon are not limited merely to the "scientific" and "technical" but extend to all "specialized" knowledge. Similarly, the expert is viewed, not in a narrow sense, but as a person qualified by "knowledge, skill, experience, training or education." Thus within the scope of the rule are not only experts in the strictest sense of the word, e.g., physicians, physicists, and architects, but also the large group sometimes called "skilled" witnesses, such as bankers or landowners testifying to land values.

COMMITTEE NOTES ON RULES—2000 AMENDMENT

Rule 702 has been amended in response to *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), and to the many cases applying *Daubert*, including *Kumho Tire Co. v. Carmichael*, 119 S.Ct. 1167 (1999). In *Daubert* the Court charged trial judges with the responsibility of acting as gatekeepers to exclude unreliable expert testimony, and the Court in *Kumho* clarified that this gatekeeper function applies to all expert testimony, not just testimony based in science. *See also Kumho*, 119 S.Ct. at 1178 (citing the Committee Note to the proposed amendment to Rule 702, which had been released for public comment before the date of the *Kumho* decision). The amendment affirms the trial court's role as gatekeeper and provides some general standards that the trial court must use to assess the reliability and helpfulness of proffered expert testimony. Consistently with *Kumho*, the Rule as amended provides that all types of expert testimony present questions of admissibility for the trial court in deciding whether the evidence is reliable and helpful. Consequently, the admissibility of all expert testimony is governed by the principles of Rule 104(a). Under that Rule, the proponent has the burden of establishing that the pertinent admissibility requirements are met by a preponderance of the evidence. *See Bourjaily v. United States*,483 U.S. 171 (1987).

Daubert set forth a non-exclusive checklist for trial courts to use in assessing the reliability of scientific expert testimony. The specific factors explicated by the *Daubert* Court are (1) whether the expert's technique or theory can be or has been tested—that is, whether the expert's theory can be challenged in some objective sense, or whether it is instead simply a subjective, conclusory approach that cannot reasonably be assessed for reliability; (2) whether the technique or theory has been subject to peer review and publication; (3) the known or potential rate of error of the technique or theory when applied; (4) the existence and maintenance of standards and controls; and (5) whether the technique or theory has been generally accepted in the scientific community. The Court in *Kumho* held that these factors might also be applicable in assessing the reliability of nonscientific expert testimony, depending upon "the particular circumstances of the particular case at issue." 119 S.Ct. at 1175.

No attempt has been made to "codify" these specific factors. *Daubert* itself emphasized that the factors were neither exclusive nor dispositive. Other cases have recognized that not all of the specific *Daubert*factors can apply to every type of expert testimony. In addition to *Kumho*, 119 S.Ct. at 1175, *see Tyus v. Urban Search Management*, 102 F.3d 256 (7th Cir. 1996) (noting that the factors mentioned by the Court in *Daubert* do not neatly apply to expert testimony from a sociologist). *See also Kannankeril v. Terminix Int'l, Inc.*, 128 F.3d 802, 809 (3d Cir. 1997) (holding that lack of peer review or publication was not dispositive where the expert's opinion was supported by "widely accepted scientific knowledge"). The standards set forth in the amendment are broad enough to require consideration of any or all of the specific *Daubert* factors where appropriate.

Courts both before and after *Daubert* have found other factors relevant in determining whether expert testimony is sufficiently reliable to be considered by the trier of fact. These factors include:

(1) Whether experts are "proposing to testify about matters growing naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying." *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311, 1317 (9th Cir. 1995).

(2) Whether the expert has unjustifiably extrapolated from an accepted premise to an unfounded conclusion. *See General Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997) (noting that in some cases a trial court "may conclude that there is simply too great an analytical gap between the data and the opinion proffered").

(3) Whether the expert has adequately accounted for obvious alternative explanations. *See Claar v. Burlington N.R.R.*, 29 F.3d 499 (9th Cir. 1994) (testimony excluded where the expert failed to consider other obvious causes for the plaintiff's condition). *Compare Ambrosini v. Labarraque*, 101 F.3d 129 (D.C.Cir. 1996) (the possibility of some uneliminated causes presents a question of weight, so long as the most obvious causes have been considered and reasonably ruled out by the expert).

(4) Whether the expert "is being as careful as he would be in his regular professional work outside his paid litigation consulting." *Sheehan v. Daily Racing Form, Inc.*, 104 F.3d 940, 942 (7th Cir. 1997). *See Kumho Tire Co. v. Carmichael*, 119 S.Ct. 1167, 1176 (1999) (*Daubert* requires the trial court to assure itself that the expert "employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field").

(5) Whether the field of expertise claimed by the expert is known to reach reliable results for the type of opinion the expert would give. *See Kumho Tire Co. v. Carmichael*, 119 S.Ct. 1167, 1175 (1999) (*Daubert's* general acceptance factor does not "help show that an expert's testimony is reliable where the discipline itself lacks reliability, as, for example, do theories grounded in any so-called generally accepted principles of astrology or necromancy."); *Moore v. Ashland Chemical, Inc.*, 151 F.3d 269 (5th Cir. 1998) (en banc) (clinical doctor was properly precluded from testifying to the toxicological cause of the plaintiff's respiratory problem, where the opinion was not sufficiently grounded in scientific methodology); *Sterling v. Velsicol Chem. Corp.*, 855 F.2d 1188 (6th Cir. 1988) (rejecting testimony based on "clinical ecology" as unfounded and unreliable).

All of these factors remain relevant to the determination of the reliability of expert testimony under the Rule as amended. Other factors may also be relevant. *See Kumho*, 119 S.Ct. 1167, 1176 ("[W]e conclude that the trial judge must have considerable leeway in deciding in a particular case how to go about determining whether particular expert testimony is reliable."). Yet no single factor is necessarily dispositive of the reliability of a particular expert's testimony. *See, e.g., Heller v. Shaw Industries, Inc.*, 167 F.3d 146, 155 (3d Cir. 1999) ("not only must each stage of the expert's testimony be reliable, but each stage must be evaluated practically and flexibly without bright-line exclusionary (or inclusionary) rules.");*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311, 1317, n.5 (9th Cir. 1995) (noting that some expert disciplines "have the courtroom as a principal theatre of operations" and as to these disciplines "the fact that the expert has developed an expertise principally for purposes of litigation will obviously not be a substantial consideration.").

A review of the caselaw after *Daubert* shows that the rejection of expert testimony is the exception rather than the rule. *Daubert* did not work a "seachange over federal evidence law," and "the trial court's role as gatekeeper is not intended to serve as a replacement for the adversary system." *United States v. 14.38 Acres of Land Situated in Leflore County, Mississippi*, 80 F.3d 1074, 1078 (5th Cir. 1996). As the Court in *Daubert* stated: "Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence." 509 U.S. at 595. Likewise, this amendment is not intended to provide an excuse for an automatic challenge to the testimony of every expert. *See Kumho Tire Co. v. Carmichael*, 119 S.Ct. 1167, 1176 (1999) (noting that the trial judge has the discretion "both to avoid unnecessary 'reliability' proceedings in ordinary cases where the reliability of an expert's methods is properly taken for granted, and to require appropriate proceedings in the less usual or more complex cases where cause for questioning the expert's reliability arises.").

When a trial court, applying this amendment, rules that an expert's testimony is reliable, this does not necessarily mean that contradictory expert testimony is unreliable. The amendment is broad enough to permit testimony that is the product of competing principles or methods in the same field of expertise.*See, e.g., Heller v. Shaw Industries, Inc.*, 167 F.3d 146, 160 (3d Cir. 1999) (expert testimony cannot be excluded simply because the expert uses one test rather than another, when both tests are accepted in the field and both reach reliable results). As the court stated in *In re Paoli R.R. Yard PCB Litigation*, 35 F.3d 717, 744 (3d Cir. 1994), proponents "do not have to demonstrate to the judge by a preponderance of the evidence that the assessments of their experts are correct, they only have to demonstrate by a preponderance of evidence that their opinions are reliable. . . . The evidentiary requirement of reliability is lower than the merits standard of correctness." *See also Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311, 1318 (9th Cir. 1995) (scientific experts might be permitted to testify if they could show that the methods they used were also employed by "a recognized minority of scientists in their field."); *Ruiz-Troche v. Pepsi Cola*, 161 F.3d 77, 85 (1st Cir. 1998) (" *Daubert* neither requires nor empowers trial courts to determine which of several competing scientific theories has the best provenance.").

The Court in *Daubert* declared that the "focus, of course, must be solely on principles and methodology, not on the conclusions they generate." 509 U.S. at 595. Yet as the Court later recognized, "conclusions and methodology are not entirely distinct from one another." *General Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997). Under the amendment, as under *Daubert*, when an expert purports to apply principles and methods in accordance with professional standards, and yet reaches a conclusion that other experts in the field would not reach, the trial court may fairly suspect that the principles and methods have not been faithfully applied. *See Lust v. Merrell Dow Pharmaceuticals, Inc.*, 89 F.3d 594, 598 (9th Cir. 1996). The amendment specifically provides that the trial court must scrutinize not only the principles and methods used by the expert, but also whether those principles and methods have been properly applied to the facts of the case. As the court noted in *In re Paoli R.R. Yard PCB Litig*., 35 F.3d 717, 745 (3d Cir. 1994), " *any*step that renders the analysis unreliable . . . renders the expert's testimony inadmissible. *This is true whether the step completely changes a reliable methodology or merely misapplies that methodology*."

If the expert purports to apply principles and methods to the facts of the case, it is important that this application be conducted reliably. Yet it might also be important in some cases for an expert to educate the factfinder about general principles, without ever attempting to apply these principles to the specific facts of the case. For example, experts might instruct the factfinder on the principles of thermodynamics, or bloodclotting, or on how financial markets respond to corporate reports, without ever knowing about or trying to tie their testimony into the facts of the case. The amendment does not alter the venerable practice of using expert testimony to educate the factfinder on general principles. For this kind of generalized testimony, Rule 702 simply requires that: (1) the expert be qualified; (2) the testimony address a subject matter on which the factfinder can be assisted by an expert; (3) the testimony be reliable; and (4) the testimony "fit" the facts of the case.

As stated earlier, the amendment does not distinguish between scientific and other forms of expert testimony. The trial court's gatekeeping function applies to testimony by any expert. *See Kumho Tire Co. v. Carmichael*, 119 S.Ct. 1167, 1171 (1999) ("We conclude that *Daubert's* general holding—setting forth the trial judge's general 'gatekeeping' obligation—applies not only to testimony based on 'scientific' knowledge, but also to testimony based on 'technical' and 'other specialized' knowledge."). While the relevant factors for determining reliability will vary from expertise to expertise, the amendment rejects the premise that an expert's testimony should be treated more permissively simply because it is outside the realm of science. An opinion from an expert who is not a scientist should receive the same degree of scrutiny for reliability as an opinion from an expert who purports to be a scientist. *See Watkins v. Telsmith, Inc*., 121 F.3d 984, 991 (5th Cir. 1997) ("[I]t seems exactly backwards that experts who purport to rely on general engineering principles and practical experience might escape screening by the district court simply by stating that their conclusions were not reached by any particular method or technique."). Some types of expert testimony will be more objectively verifiable, and subject to the expectations of falsifiability, peer review, and publication, than others. Some types of expert testimony will not rely on anything like a scientific method, and so will have to be evaluated by reference to other standard principles attendant to the particular area of expertise. The trial judge in all cases of proffered expert testimony must find that it is properly grounded, well-reasoned, and not speculative before it can be admitted. The expert's testimony must be grounded in an accepted body of learning or experience in the expert's field, and the expert must explain how the conclusion is so grounded. *See, e.g*., American College of Trial Lawyers, *Standards and Procedures for Determining the Admissibility of Expert Testimony after Daubert*, 157 F.R.D. 571, 579 (1994) ("[W]hether the testimony concerns economic principles, accounting standards, property valuation or other non-scientific subjects, it should be evaluated by reference to the 'knowledge and experience' of that particular field.").

The amendment requires that the testimony must be the product of reliable principles and methods that are reliably applied to the facts of the case. While the terms "principles" and "methods" may convey a certain impression when applied to scientific knowledge, they remain relevant when applied to testimony based on technical or other specialized knowledge. For example, when a law enforcement agent testifies regarding the use of code words in a drug transaction, the principle used by the agent is that participants in such transactions regularly use code words to conceal the nature of their activities. The method used by the agent is the application of extensive experience to analyze the meaning of the conversations. So long as the principles and methods are reliable and applied reliably to the facts of the case, this type of testimony should be admitted.

Nothing in this amendment is intended to suggest that experience alone—or experience in conjunction with other knowledge, skill, training or education—may not provide a sufficient foundation for expert testimony. To the contrary, the text of Rule 702 expressly contemplates that an expert may be qualified on the basis of experience. In certain fields, experience is the predominant, if not sole, basis for a great deal of reliable expert testimony. *See, e.g., United States v. Jones*, 107 F.3d 1147 (6th Cir. 1997) (no abuse of discretion in admitting the testimony of a handwriting examiner who had years of practical experience and extensive training, and who explained his methodology in detail); *Tassin v. Sears Roebuck*, 946 F.Supp. 1241, 1248 (M.D.La. 1996) (design engineer's testimony can be admissible when the expert's opinions "are based on facts, a reasonable investigation, and traditional technical/mechanical expertise, and he provides a reasonable link between the information and procedures he uses and the conclusions he reaches"). *See also Kumho Tire Co. v. Carmichael*, 119 S.Ct. 1167, 1178 (1999) (stating that "no one denies that an expert might draw a conclusion from a set of observations based on extensive and specialized experience.").

If the witness is relying solely or primarily on experience, then the witness must explain how that experience leads to the conclusion reached, why that experience is a sufficient basis for the opinion, and how that experience is reliably applied to the facts. The trial court's gatekeeping function requires more than simply "taking the expert's word for it." *See Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311, 1319 (9th Cir. 1995) ("We've been presented with only the experts' qualifications, their conclusions and their assurances of reliability. Under *Daubert*, that's not enough."). The more subjective and controversial the expert's inquiry, the more likely the testimony should be excluded as unreliable. *See O'Conner v. Commonwealth Edison Co.*, 13 F.3d 1090 (7th Cir. 1994) (expert testimony based on a completely subjective methodology held properly excluded). *See also Kumho Tire Co. v. Carmichael*, 119 S.Ct. 1167, 1176 (1999) ("[I]t will at times be useful to ask even of a witness whose expertise is based purely on experience, say, a perfume tester able to distinguish among 140 odors at a sniff, whether his preparation is of a kind that others in the field would recognize as acceptable.").

Subpart (1) of Rule 702 calls for a quantitative rather than qualitative analysis. The amendment requires that expert testimony be based on sufficient underlying "facts or data." The term "data" is intended to encompass the reliable opinions of other experts. See the original Advisory Committee Note to Rule 703. The language "facts or data" is broad enough to allow an expert to rely on hypothetical facts that are supported by the evidence. *Id.*

When facts are in dispute, experts sometimes reach different conclusions based on competing versions of the facts. The emphasis in the amendment on "sufficient facts or data" is not intended to authorize a trial court to exclude an expert's testimony on the ground that the court believes one version of the facts and not the other.

There has been some confusion over the relationship between Rules 702 and 703. The amendment makes clear that the sufficiency of the basis of an expert's testimony is to be decided under Rule 702. Rule 702 sets forth the overarching requirement of reliability, and an analysis of the sufficiency of the expert's basis cannot be divorced from the ultimate reliability of the expert's opinion. In contrast, the "reasonable reliance" requirement of Rule 703 is a relatively narrow inquiry. When an expert relies on inadmissible information, Rule 703 requires the trial court to determine whether that information is of a type reasonably relied on by other experts in the field. If so, the expert can rely on the information in reaching an opinion. However, the question whether the expert is relying on a *sufficient* basis of information—whether admissible information or not—is governed by the requirements of Rule 702.

The amendment makes no attempt to set forth procedural requirements for exercising the trial court's gatekeeping function over expert testimony. *See* Daniel J. Capra, *The Daubert Puzzle*, 38 Ga.L.Rev. 699, 766 (1998) ("Trial courts should be allowed substantial discretion in dealing with *Daubert* questions; any attempt to codify procedures will likely give rise to unnecessary changes in practice and create difficult questions for appellate review."). Courts have shown considerable ingenuity and flexibility in considering challenges to expert testimony under *Daubert*, and it is contemplated that this will continue under the amended Rule. *See, e.g., Cortes-Irizarry v. Corporacion Insular*, 111 F.3d 184 (1st Cir. 1997) (discussing the application of *Daubert* in ruling on a motion for summary judgment); *In re Paoli R.R. Yard PCB Litig.*,35 F.3d 717, 736, 739 (3d Cir. 1994) (discussing the use of *in limine* hearings); *Claar v. Burlington N.R.R.*, 29 F.3d 499, 502–05 (9th Cir. 1994) (discussing the trial court's technique of ordering experts to submit serial affidavits explaining the reasoning and methods underlying their conclusions).

The amendment continues the practice of the original Rule in referring to a qualified witness as an "expert." This was done to provide continuity and to minimize change. The use of the term "expert" in the Rule does not, however, mean that a jury should actually be informed that a qualified witness is testifying as an "expert." Indeed, there is much to be said for a practice that prohibits the use of the term "expert" by both the parties and the court at trial. Such a practice "ensures that trial courts do not inadvertently put their stamp of authority" on a witness's opinion, and protects against the jury's being "overwhelmed by the so-called 'experts'." Hon. Charles Richey, *Proposals to Eliminate the Prejudicial Effect of the Use of the Word "Expert" Under the Federal Rules of Evidence in Criminal and Civil Jury Trials*, 154 F.R.D. 537, 559 (1994) (setting forth limiting instructions and a standing order employed to prohibit the use of the term "expert" in jury trials).

GAP REPORT—PROPOSED AMENDMENT TO RULE 702.

The Committee made the following changes to the published draft of the proposed amendment to Evidence Rule 702:

1. The word "reliable" was deleted from Subpart (1) of the proposed amendment, in order to avoid an overlap with Evidence Rule 703, and to clarify that an expert opinion need not be

excluded simply because it is based on hypothetical facts. The Committee Note was amended to accord with this textual change.

2. The Committee Note was amended throughout to include pertinent references to the Supreme Court's decision in *Kumho Tire Co. v. Carmichael*, which was rendered after the proposed amendment was released for public comment. Other citations were updated as well.

3. The Committee Note was revised to emphasize that the amendment is not intended to limit the right to jury trial, nor to permit a challenge to the testimony of every expert, nor to preclude the testimony of experience-based experts, nor to prohibit testimony based on competing methodologies within a field of expertise.

4. Language was added to the Committee Note to clarify that no single factor is necessarily dispositive of the reliability inquiry mandated by Evidence Rule 702.

Committee Notes on Rules—2011 Amendment

The language of Rule 702 has been amended as part of the restyling of the Evidence Rules to make them more easily understood and to make style and terminology consistent throughout the rules. These changes are intended to be stylistic only. There is no intent to change any result in any ruling on evidence admissibility.

FRYE v. UNITED STATES

54 App. D. C. 46, 293 F. 1013

No. 3968

Court of Appeals of District of Columbia

Submitted November 7, 1923 December 3, 1923, Decided

Before SMYTH, Chief Justice, VAN ORSDEL, Associate Justice, and MARTIN, Presiding Judge of the United States Court of Customs Appeals.

VAN ORSDEL, Associate Justice. Appellant, defendant below, was convicted of the crime of murder in the second degree, and from the judgment prosecutes this appeal.

A single assignment of error is presented for our consideration. In the course of the trial counsel for defendant offered an expert witness to testify to the result of a deception test made upon defendant. The test is described as the systolic blood pressure deception test. It is asserted that blood pressure is influenced by change in the emotions of the witness, and that the systolic blood pressure rises are brought about by nervous impulses sent to the sympathetic branch of the autonomic nervous system. Scientific experiments, it is claimed, have demonstrated that fear, rage, and pain always produce a rise of systolic blood pressure, and that conscious deception or falsehood, concealment of facts, or guilt of crime, accompanied by fear of detection when the person is under examination, raises the systolic blood pressure in a curve, which corresponds exactly to the struggle going on in the subject's mind, between fear and attempted control of that fear, as the examination touches the vital points in respect of which he is attempting to deceive the examiner.

In other words, the theory seems to be that truth is spontaneous, and comes without conscious effort, while the utterance of a falsehood requires a conscious effort, which is reflected in the blood pressure. The rise thus produced is easily detected and distinguished from the rise produced by mere fear of the examination itself. In the former instance, the pressure rises higher than in the latter, and is more pronounced as the examination proceeds, while in the latter case, if the subject is telling the truth, the pressure registers highest at the beginning of the examination, and gradually diminishes as the examination proceeds.

Prior to the trial defendant was subjected to this deception test, and counsel offered the scientist who conducted the test as an expert to testify to the results obtained. The offer was objected to by counsel for the government, and the court sustained the objection. Counsel for defendant then offered to have the proffered witness conduct a test in the presence of the jury. This also was denied.

Counsel for defendant, in their able presentation of the novel question involved, correctly state in their brief that no cases directly in point have been found. The broad ground, however, upon which they plant their case, is succinctly stated in their brief as follows:

"The rule is that the opinions of experts or skilled witnesses are admissible in evidence in those cases in which the matter of inquiry is such that inexperienced persons are unlikely to prove capable of forming a correct judgment upon it, for the reason that the subject-matter so far partakes of a science, art, or trade as to require a previous habit or experience or study in it, in order to acquire a knowledge of it. When the question involved does not lie within the range of common experience or common knowledge, but requires special experience or special knowledge, then the opinions of witnesses skilled in that particular science, art, or trade to which the question relates are admissible in evidence."

Numerous cases are cited in support of this rule. Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.

We think the systolic blood pressure deception test has not yet gained such standing and scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made.

The judgment is affirmed.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S. Ct. 2786 (1993). Document in public domain.

Syllabus

DAUBERT ET UX., INDIVIDUALLY AND AS GUARDIANS AD LITEM FOR DAUBERT, ET AL. *v.* MERRELL DOW PHARMACEUTICALS, INC.

CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT

No. 92–102. Argued March 30, 1993—Decided June 28, 1993

Petitioners, two minor children and their parents, alleged in their suit against respondent that the children's serious birth defects had been caused by the mothers' prenatal ingestion of Bendectin, a prescription drug marketed by respondent. The District Court granted respondent summary judgment based on a well-credentialed expert's affidavit concluding, upon reviewing the extensive published scientific literature on the subject, that maternal use of Bendectin has not been shown to be a risk factor for human birth defects. Although petitioners had responded with the testimony of eight other well-credentialed experts, who based their conclusion that Bendectin can cause birth defects on animal studies, chemical structure analyses, and the unpublished "reanalysis" of previously published human statistical studies, the court determined that this evidence did not meet the applicable "general acceptance" standard for the admission of expert testimony. The Court of Appeals agreed and affirmed, citing *Frye* v. *United States,* 54 App. D. C. 46, 47, 293 F. 1013, 1014, for the rule that expert opinion based on a scientific technique is inadmissible unless the technique is "generally accepted" as reliable in the relevant scientific community.

Held: The Federal Rules of Evidence, not *Frye,* provide the standard for admitting expert scientific testimony in a federal trial. Pp. 585–597.

(a) *Frye*'s "general acceptance" test was superseded by the Rules' adoption. The Rules occupy the field, *United States* v. *Abel,* 469 U. S. 45, 49, and, although the common law of evidence may serve as an aid to their application, *id.,* at 51–52, respondent's assertion that they somehow assimilated *Frye* is unconvincing. Nothing in the Rules as a whole or in the text and drafting history of Rule 702, which specifically governs expert testimony, gives any indication that "general acceptance" is a necessary precondition to the admissibility of scientific evidence. Moreover, such a rigid standard would be at odds with the Rules' liberal thrust and their general approach of relaxing the traditional barriers to "opinion" testimony. Pp. 585–589.

(b) The Rules—especially Rule 702—place appropriate limits on the admissibility of purportedly scientific evidence by assigning to the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand. The reliability standard is established by Rule 702's requirement that an expert's testimony pertain to "scientific . . . knowledge," since the adjective "scientific" implies a grounding in science's methods and procedures, while the word "knowledge" connotes a body of known facts or of ideas inferred from such facts or accepted as true on good grounds. The Rule's requirement that the testimony "assist the trier of fact to understand the evidence or to determine a fact in issue" goes primarily to relevance by demanding a valid scientific connection to the pertinent inquiry as a precondition to admissibility. Pp. 589–592.

(c) Faced with a proffer of expert scientific testimony under Rule 702, the trial judge, pursuant to Rule 104(a), must make a preliminary assessment of whether the testimony's underlying reasoning or methodology is scientifically valid and properly can be applied to the facts at issue. Many considerations will bear on the inquiry, including whether the theory or technique in question can be (and has been) tested, whether it has been subjected to peer review and publication, its known or potential error rate and the existence and maintenance of standards controlling its operation, and whether it has attracted widespread acceptance within a relevant scientific community. The inquiry is a flexible one, and its focus must be solely on principles and methodology, not on the conclusions that they generate. Throughout, the judge should also be mindful of other applicable Rules. Pp. 592–595.

(d) Cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof, rather than wholesale exclusion under an uncompromising "general acceptance" standard, is the appropriate means by which evidence based on valid principles may be challenged. That even limited screening by the trial judge, on occasion, will prevent the jury from hearing of authentic scientific breakthroughs is simply a consequence of the fact that the Rules are not designed to seek cosmic understanding but, rather, to resolve legal disputes. Pp. 595–597.

951 F. 2d 1128, vacated and remanded.

Blackmun, J., delivered the opinion for a unanimous Court with respect to Parts I and II–A, and the opinion of the Court with respect to Parts II–B, II–C, III, and IV, in which White, O'Connor, Scalia, Kennedy, Souter, and Thomas, JJ., joined. Rehnquist, C. J., filed an opinion concurring in part and dissenting in part, in which Stevens, J., joined, *post,* p. 598.

Michael H. Gottesman argued the cause for petitioners. With him on the briefs were *Kenneth J. Chesebro*, *Barry J. Nace*, *David L. Shapiro*, and *Mary G. Gillick*.

Charles Fried argued the cause for respondent. With him on the brief were *Charles R. Nesson*, *Joel I. Klein*, *Richard G. Taranto*, *Hall R. Marston*, *George E. Berry*, *Edward H. Stratemeier*, and *W. Glenn Forrester*.*

---

*Briefs of *amici curiae* urging reversal were filed for the State of Texas et al. by *Dan Morales*, Attorney General of Texas, *Mark Barnett*, Attorney General of South Dakota, *Marc Racicot*, Attorney General of Montana, *Larry EchoHawk*, Attorney General of Idaho, and *Brian Stuart Koukoutchos;* for the American Society of Law, Medicine and Ethics et al. by *Joan E. Bertin*, *Marsha S. Berzon*, and *Albert H. Meyerhoff;* for the Association of Trial Lawyers of America by *Jeffrey Robert White* and *Roxanne Barton Conlin;* for Ronald Bayer et al. by *Brian Stuart Koukoutchos*, *Priscilla Budeiri*, *Arthur Bryant*, and *George W. Conk;* and for Daryl E. Chubin et al. by *Ron Simon* and *Nicole Schultheis*.

Briefs of *amici curiae* urging affirmance were filed for the United States by *Acting Solicitor General Wallace*, *Assistant Attorney General Gerson*, *Miguel A. Estrada*, *Michael Jay Singer*, and *John P. Schnitker;* for the American Insurance Association by *William J. Kilberg*, *Paul Blankenstein*, *Bradford R. Clark*, and *Craig A. Berrington;* for the American Medical Association et al. by *Carter G. Phillips*, *Mark D. Hopson*, and *Jack R. Bierig;* for the American Tort Reform Association by *John G. Kester* and *John W. Vardaman, Jr.;* for the Chamber of Commerce of the United States by *Timothy B. Dyk*, *Stephen A. Bokat*, and *Robin S. Conrad;* for the Pharmaceutical Manufacturers Association by *Louis R. Cohen* and *Daniel Marcus;* for the Product Liability Advisory Council, Inc., et al. by *Victor E. Schwartz*, *Robert P. Charrow*, and *Paul F. Rothstein;* for the Washington Legal Foundation by *Scott G. Campbell*, *Daniel J. Popeo*, and *Richard A. Samp;* and for Nicolaas Bloembergen et al. by *Martin S. Kaufman*.

Briefs of *amici curiae* were filed for the American Association for the Advancement of Science et al. by *Richard A. Meserve* and *Bert Black;* for the American College of Legal Medicine by *Miles J. Zaremski;* for the Carnegie Commission on Science, Technology, and Government by *Steven G. Gallagher*, *Elizabeth H. Esty*, and *Margaret A. Berger;* for the Defense Research Institute, Inc., by *Joseph A. Sherman*, *E. Wayne Taff*, and *Harvey L. Kaplan;* for the New England Journal of Medicine et al. by *Michael Malina* and *Jeffrey I. D. Lewis;* for A Group of American Law Professors

JUSTICE BLACKMUN delivered the opinion of the Court.

In this case we are called upon to determine the standard for admitting expert scientific testimony in a federal trial.

I

Petitioners Jason Daubert and Eric Schuller are minor children born with serious birth defects. They and their parents sued respondent in California state court, alleging that the birth defects had been caused by the mothers' ingestion of Bendectin, a prescription antinausea drug marketed by respondent. Respondent removed the suits to federal court on diversity grounds.

After extensive discovery, respondent moved for summary judgment, contending that Bendectin does not cause birth defects in humans and that petitioners would be unable to come forward with any admissible evidence that it does. In support of its motion, respondent submitted an affidavit of Steven H. Lamm, physician and epidemiologist, who is a well-credentialed expert on the risks from exposure to various chemical substances.[1] Doctor Lamm stated that he had reviewed all the literature on Bendectin and human birth defects—more than 30 published studies involving over 130,000 patients. No study had found Bendectin to be a human teratogen (*i. e.*, a substance capable of causing malformations in fetuses). On the basis of this review, Doctor Lamm concluded that maternal use of Bendectin during the first trimester of pregnancy has not been shown to be a risk factor for human birth defects.

---

by *Donald N. Bersoff;* for Alvan R. Feinstein by *Don M. Kennedy, Loretta M. Smith,* and *Richard A. Oetheimer;* and for Kenneth Rothman et al. by *Neil B. Cohen.*

[1] Doctor Lamm received his master's and doctor of medicine degrees from the University of Southern California. He has served as a consultant in birth-defect epidemiology for the National Center for Health Statistics and has published numerous articles on the magnitude of risk from exposure to various chemical and biological substances. App. 34–44.

Petitioners did not (and do not) contest this characterization of the published record regarding Bendectin. Instead, they responded to respondent's motion with the testimony of eight experts of their own, each of whom also possessed impressive credentials.[2] These experts had concluded that Bendectin can cause birth defects. Their conclusions were based upon "in vitro" (test tube) and "in vivo" (live) animal studies that found a link between Bendectin and malformations; pharmacological studies of the chemical structure of Bendectin that purported to show similarities between the structure of the drug and that of other substances known to cause birth defects; and the "reanalysis" of previously published epidemiological (human statistical) studies.

The District Court granted respondent's motion for summary judgment. The court stated that scientific evidence is admissible only if the principle upon which it is based is " 'sufficiently established to have general acceptance in the field to which it belongs.' " 727 F. Supp. 570, 572 (SD Cal. 1989), quoting *United States* v. *Kilgus*, 571 F. 2d 508, 510 (CA9 1978). The court concluded that petitioners' evidence did not meet this standard. Given the vast body of epidemiological data concerning Bendectin, the court held, expert opinion which is not based on epidemiological evidence

---

[2] For example, Shanna Helen Swan, who received a master's degree in biostatistics from Columbia University and a doctorate in statistics from the University of California at Berkeley, is chief of the section of the California Department of Health and Services that determines causes of birth defects and has served as a consultant to the World Health Organization, the Food and Drug Administration, and the National Institutes of Health. *Id.*, at 113–114, 131–132. Stuart A. Newman, who received his bachelor's degree in chemistry from Columbia University and his master's and doctorate in chemistry from the University of Chicago, is a professor at New York Medical College and has spent over a decade studying the effect of chemicals on limb development. *Id.*, at 54–56. The credentials of the others are similarly impressive. See *id.*, at 61–66, 73–80, 148–153, 187–192, and Attachments 12, 20, 21, 26, 31, and 32 to Petitioners' Opposition to Summary Judgment in No. 84–2013–G(I) (SD Cal.).

is not admissible to establish causation. 727 F. Supp., at 575. Thus, the animal-cell studies, live-animal studies, and chemical-structure analyses on which petitioners had relied could not raise by themselves a reasonably disputable jury issue regarding causation. *Ibid.* Petitioners' epidemiological analyses, based as they were on recalculations of data in previously published studies that had found no causal link between the drug and birth defects, were ruled to be inadmissible because they had not been published or subjected to peer review. *Ibid.*

The United States Court of Appeals for the Ninth Circuit affirmed. 951 F. 2d 1128 (1991). Citing *Frye* v. *United States,* 54 App. D. C. 46, 47, 293 F. 1013, 1014 (1923), the court stated that expert opinion based on a scientific technique is inadmissible unless the technique is "generally accepted" as reliable in the relevant scientific community. 951 F. 2d, at 1129–1130. The court declared that expert opinion based on a methodology that diverges "significantly from the procedures accepted by recognized authorities in the field . . . cannot be shown to be 'generally accepted as a reliable technique.'" *Id.*, at 1130, quoting *United States* v. *Solomon*, 753 F. 2d 1522, 1526 (CA9 1985).

The court emphasized that other Courts of Appeals considering the risks of Bendectin had refused to admit reanalyses of epidemiological studies that had been neither published nor subjected to peer review. 951 F. 2d, at 1130–1131. Those courts had found unpublished reanalyses "particularly problematic in light of the massive weight of the original published studies supporting [respondent's] position, all of which had undergone full scrutiny from the scientific community." *Id.*, at 1130. Contending that reanalysis is generally accepted by the scientific community only when it is subjected to verification and scrutiny by others in the field, the Court of Appeals rejected petitioners' reanalyses as "unpublished, not subjected to the normal peer review process and generated solely for use in litigation." *Id.*, at 1131. The court concluded that petitioners' evidence provided an insufficient foundation to allow admission of expert testimony that Bendectin caused their injuries and, accordingly, that petitioners could not satisfy their burden of proving causation at trial.

We granted certiorari, 506 U. S. 914 (1992), in light of sharp divisions among the courts regarding the proper standard for the admission of expert testimony. Compare, *e. g., United States* v. *Shorter*, 257 U. S. App. D. C. 358, 363–364, 809 F. 2d 54, 59–60 (applying the "general acceptance" standard), cert. denied, 484 U. S. 817 (1987), with *DeLuca* v. *Merrell Dow Pharmaceuticals, Inc.*, 911 F. 2d 941, 955 (CA3 1990) (rejecting the "general acceptance" standard).

II

A

In the 70 years since its formulation in the *Frye* case, the "general acceptance" test has been the dominant standard for determining the admissibility of novel scientific evidence at trial. See E. Green & C. Nesson, Problems, Cases, and Materials on Evidence 649 (1983). Although under increasing attack of late, the rule continues to be followed by a majority of courts, including the Ninth Circuit.[3]

The *Frye* test has its origin in a short and citation-free 1923 decision concerning the admissibility of evidence derived from a systolic blood pressure deception test, a crude precursor to the polygraph machine. In what has become a famous (perhaps infamous) passage, the then Court of Appeals for the District of Columbia described the device and its operation and declared:

> "Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages

---

[3] For a catalog of the many cases on either side of this controversy, see P. Giannelli & E. Imwinkelried, Scientific Evidence § 1–5, pp. 10–14 (1986 and Supp. 1991).

> is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, *the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.*" 54 App. D. C., at 47, 293 F., at 1014 (emphasis added).

Because the deception test had "not yet gained such standing and scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made," evidence of its results was ruled inadmissible. *Ibid.*

The merits of the *Frye* test have been much debated, and scholarship on its proper scope and application is legion.[4]

[4] See, *e. g.*, Green, Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of *Agent Orange* and Bendectin Litigation, 86 Nw. U. L. Rev. 643 (1992) (hereinafter Green); Becker & Orenstein, The Federal Rules of Evidence After Sixteen Years—The Effect of "Plain Meaning" Jurisprudence, the Need for an Advisory Committee on the Rules of Evidence, and Suggestions for Selective Revision of the Rules, 60 Geo. Wash. L. Rev. 857, 876–885 (1992); Hanson, James Alphonzo Frye is Sixty-Five Years Old; Should He Retire?, 16 West. St. U. L. Rev. 357 (1989); Black, A Unified Theory of Scientific Evidence, 56 Ford. L. Rev. 595 (1988); Imwinkelried, The "Bases" of Expert Testimony: The Syllogistic Structure of Scientific Testimony, 67 N. C. L. Rev. 1 (1988); Proposals for a Model Rule on the Admissibility of Scientific Evidence, 26 Jurimetrics J. 235 (1986); Giannelli, The Admissibility of Novel Scientific Evidence: *Frye* v. *United States*, a Half-Century Later, 80 Colum. L. Rev. 1197 (1980); The Supreme Court, 1986 Term, 101 Harv. L. Rev. 7, 119, 125–127 (1987).

Indeed, the debates over *Frye* are such a well-established part of the academic landscape that a distinct term—"*Frye*-ologist"—has been advanced to describe those who take part. See Behringer, Introduction, Proposals for a Model Rule on the Admissibility of Scientific Evidence, 26 Jurimetrics J. 237, 239 (1986), quoting Lacey, Scientific Evidence, 24 Jurimetrics J. 254, 264 (1984).

Petitioners' primary attack, however, is not on the content but on the continuing authority of the rule. They contend that the *Frye* test was superseded by the adoption of the Federal Rules of Evidence.[5] We agree.

We interpret the legislatively enacted Federal Rules of Evidence as we would any statute. *Beech Aircraft Corp.* v. *Rainey*, 488 U. S. 153, 163 (1988). Rule 402 provides the baseline:

> "All relevant evidence is admissible, except as otherwise provided by the Constitution of the United States, by Act of Congress, by these rules, or by other rules prescribed by the Supreme Court pursuant to statutory authority. Evidence which is not relevant is not admissible."

"Relevant evidence" is defined as that which has "any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." Rule 401. The Rules' basic standard of relevance thus is a liberal one.

Frye, of course, predated the Rules by half a century. In *United States* v. *Abel*, 469 U. S. 45 (1984), we considered the pertinence of background common law in interpreting the Rules of Evidence. We noted that the Rules occupy the field, *id.*, at 49, but, quoting Professor Cleary, the Reporter,

---

[5] Like the question of *Frye*'s merit, the dispute over its survival has divided courts and commentators. Compare, *e. g.*, *United States* v. *Williams*, 583 F. 2d 1194 (CA2 1978) (*Frye* is superseded by the Rules of Evidence), cert. denied, 439 U. S. 1117 (1979), with *Christophersen* v. *Allied-Signal Corp.*, 939 F. 2d 1106, 1111, 1115–1116 (CA5 1991) (en banc) (*Frye* and the Rules coexist), cert. denied, 503 U. S. 912 (1992), 3 J. Weinstein & M. Berger, Weinstein's Evidence ¶ 702[03], pp. 702–36 to 702–37 (1988) (hereinafter Weinstein & Berger) (*Frye* is dead), and M. Graham, Handbook of Federal Evidence § 703.2 (3d ed. 1991) (*Frye* lives). See generally P. Giannelli & E. Imwinkelried, Scientific Evidence § 1–5, at 28–29 (citing authorities).

explained that the common law nevertheless could serve as an aid to their application:

> " 'In principle, under the Federal Rules no common law of evidence remains. "All relevant evidence is admissible, except as otherwise provided . . . ." In reality, of course, the body of common law knowledge continues to exist, though in the somewhat altered form of a source of guidance in the exercise of delegated powers.' " *Id.*, at 51–52.

We found the common-law precept at issue in the *Abel* case entirely consistent with Rule 402's general requirement of admissibility, and considered it unlikely that the drafters had intended to change the rule. *Id.*, at 50–51. In *Bourjaily* v. *United States*, 483 U. S. 171 (1987), on the other hand, the Court was unable to find a particular common-law doctrine in the Rules, and so held it superseded.

Here there is a specific Rule that speaks to the contested issue. Rule 702, governing expert testimony, provides:

> "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise."

Nothing in the text of this Rule establishes "general acceptance" as an absolute prerequisite to admissibility. Nor does respondent present any clear indication that Rule 702 or the Rules as a whole were intended to incorporate a "general acceptance" standard. The drafting history makes no mention of *Frye*, and a rigid "general acceptance" requirement would be at odds with the "liberal thrust" of the Federal Rules and their "general approach of relaxing the traditional barriers to 'opinion' testimony." *Beech Aircraft Corp.* v. *Rainey*, 488 U. S., at 169 (citing Rules 701 to 705). See also Weinstein, Rule 702 of the Federal Rules of Evidence is

Sound; It Should Not Be Amended, 138 F. R. D. 631 (1991) ("The Rules were designed to depend primarily upon lawyer-adversaries and sensible triers of fact to evaluate conflicts"). Given the Rules' permissive backdrop and their inclusion of a specific rule on expert testimony that does not mention "general acceptance," the assertion that the Rules somehow assimilated *Frye* is unconvincing. *Frye* made "general acceptance" the exclusive test for admitting expert scientific testimony. That austere standard, absent from, and incompatible with, the Federal Rules of Evidence, should not be applied in federal trials.[6]

B

That the *Frye* test was displaced by the Rules of Evidence does not mean, however, that the Rules themselves place no limits on the admissibility of purportedly scientific evidence.[7] Nor is the trial judge disabled from screening such evidence. To the contrary, under the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.

The primary locus of this obligation is Rule 702, which clearly contemplates some degree of regulation of the subjects and theories about which an expert may testify. "*If scientific*, technical, or other specialized *knowledge will assist the trier of fact* to understand the evidence or to determine a fact in issue" an expert "may testify *thereto.*" (Emphasis added.) The subject of an expert's testimony must

---

[6] Because we hold that *Frye* has been superseded and base the discussion that follows on the content of the congressionally enacted Federal Rules of Evidence, we do not address petitioners' argument that application of the *Frye* rule in this diversity case, as the application of a judge-made rule affecting substantive rights, would violate the doctrine of *Erie R. Co.* v. *Tompkins*, 304 U. S. 64 (1938).

[7] The Chief Justice "do[es] not doubt that Rule 702 confides to the judge some gatekeeping responsibility," *post*, at 600, but would neither say how it does so nor explain what that role entails. We believe the better course is to note the nature and source of the duty.

be "scientific . . . knowledge."[8] The adjective "scientific" implies a grounding in the methods and procedures of science. Similarly, the word "knowledge" connotes more than subjective belief or unsupported speculation. The term "applies to any body of known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds." Webster's Third New International Dictionary 1252 (1986). Of course, it would be unreasonable to conclude that the subject of scientific testimony must be "known" to a certainty; arguably, there are no certainties in science. See, *e. g.,* Brief for Nicolaas Bloembergen et al. as *Amici Curiae* 9 ("Indeed, scientists do not assert that they know what is immutably 'true'—they are committed to searching for new, temporary, theories to explain, as best they can, phenomena"); Brief for American Association for the Advancement of Science et al. as *Amici Curiae* 7–8 ("Science is not an encyclopedic body of knowledge about the universe. Instead, it represents a *process* for proposing and refining theoretical explanations about the world that are subject to further testing and refinement" (emphasis in original)). But, in order to qualify as "scientific knowledge," an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation—*i. e.,* "good grounds," based on what is known. In short, the requirement that an expert's testimony pertain to "scientific knowledge" establishes a standard of evidentiary reliability.[9]

---

[8] Rule 702 also applies to "technical, or other specialized knowledge." Our discussion is limited to the scientific context because that is the nature of the expertise offered here.

[9] We note that scientists typically distinguish between "validity" (does the principle support what it purports to show?) and "reliability" (does application of the principle produce consistent results?). See Black, 56 Ford. L. Rev., at 599. Although "the difference between accuracy, validity, and reliability may be such that each is distinct from the other by no more than a hen's kick," Starrs, *Frye v. United States* Restructured and Revitalized: A Proposal to Amend Federal Evidence Rule 702, 26 Jurimet-

Rule 702 further requires that the evidence or testimony "assist the trier of fact to understand the evidence or to determine a fact in issue." This condition goes primarily to relevance. "Expert testimony which does not relate to any issue in the case is not relevant and, ergo, non-helpful." 3 Weinstein & Berger ¶ 702[02], p. 702–18. See also *United States* v. *Downing*, 753 F. 2d 1224, 1242 (CA3 1985) ("An additional consideration under Rule 702—and another aspect of relevancy—is whether expert testimony proffered in the case is sufficiently tied to the facts of the case that it will aid the jury in resolving a factual dispute"). The consideration has been aptly described by Judge Becker as one of "fit." *Ibid.* "Fit" is not always obvious, and scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes. See Starrs, *Frye v. United States* Restructured and Revitalized: A Proposal to Amend Federal Evidence Rule 702, 26 Jurimetrics J. 249, 258 (1986). The study of the phases of the moon, for example, may provide valid scientific "knowledge" about whether a certain night was dark, and if darkness is a fact in issue, the knowledge will assist the trier of fact. However (absent creditable grounds supporting such a link), evidence that the moon was full on a certain night will not assist the trier of fact in determining whether an individual was unusually likely to have behaved irrationally on that night. Rule 702's "helpfulness"

---

rics J. 249, 256 (1986), our reference here is to *evidentiary* reliability—that is, trustworthiness. Cf., *e. g.,* Advisory Committee's Notes on Fed. Rule Evid. 602, 28 U. S. C. App., p. 755 ("'[T]he rule requiring that a witness who testifies to a fact which can be perceived by the senses must have had an opportunity to observe, and must have actually observed the fact' is a 'most pervasive manifestation' of the common law insistence upon 'the most reliable sources of information'" (citation omitted)); Advisory Committee's Notes on Art. VIII of Rules of Evidence, 28 U. S. C. App., p. 770 (hearsay exceptions will be recognized only "under circumstances supposed to furnish guarantees of trustworthiness"). In a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity.*

standard requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility.

That these requirements are embodied in Rule 702 is not surprising. Unlike an ordinary witness, see Rule 701, an expert is permitted wide latitude to offer opinions, including those that are not based on firsthand knowledge or observation. See Rules 702 and 703. Presumably, this relaxation of the usual requirement of firsthand knowledge—a rule which represents "a 'most pervasive manifestation' of the common law insistence upon 'the most reliable sources of information,' " Advisory Committee's Notes on Fed. Rule Evid. 602, 28 U. S. C. App., p. 755 (citation omitted)—is premised on an assumption that the expert's opinion will have a reliable basis in the knowledge and experience of his discipline.

C

Faced with a proffer of expert scientific testimony, then, the trial judge must determine at the outset, pursuant to Rule 104(a),[10] whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact to understand or determine a fact in issue.[11] This entails a preliminary assessment of whether the reasoning or method-

[10] Rule 104(a) provides:

"Preliminary questions concerning the qualification of a person to be a witness, the existence of a privilege, or the admissibility of evidence shall be determined by the court, subject to the provisions of subdivision (b) [pertaining to conditional admissions]. In making its determination it is not bound by the rules of evidence except those with respect to privileges." These matters should be established by a preponderance of proof. See *Bourjaily* v. *United States*, 483 U. S. 171, 175–176 (1987).

[11] Although the *Frye* decision itself focused exclusively on "novel" scientific techniques, we do not read the requirements of Rule 702 to apply specially or exclusively to unconventional evidence. Of course, well-established propositions are less likely to be challenged than those that are novel, and they are more handily defended. Indeed, theories that are so firmly established as to have attained the status of scientific law, such as the laws of thermodynamics, properly are subject to judicial notice under Federal Rule of Evidence 201.

ology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue. We are confident that federal judges possess the capacity to undertake this review. Many factors will bear on the inquiry, and we do not presume to set out a definitive checklist or test. But some general observations are appropriate.

Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested. "Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human inquiry." Green 645. See also C. Hempel, Philosophy of Natural Science 49 (1966) ("[T]he statements constituting a scientific explanation must be capable of empirical test"); K. Popper, Conjectures and Refutations: The Growth of Scientific Knowledge 37 (5th ed. 1989) ("[T]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability") (emphasis deleted).

Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication. Publication (which is but one element of peer review) is not a *sine qua non* of admissibility; it does not necessarily correlate with reliability, see S. Jasanoff, The Fifth Branch: Science Advisors as Policymakers 61–76 (1990), and in some instances well-grounded but innovative theories will not have been published, see Horrobin, The Philosophical Basis of Peer Review and the Suppression of Innovation, 263 JAMA 1438 (1990). Some propositions, moreover, are too particular, too new, or of too limited interest to be published. But submission to the scrutiny of the scientific community is a component of "good science," in part because it increases the likelihood that substantive flaws in methodology will be detected. See J. Ziman, Reliable Knowledge: An Exploration of the Grounds for Belief in Science 130–133 (1978); Relman & Angell, How Good Is Peer Review?, 321 New Eng. J. Med. 827 (1989). The fact of publication (or lack thereof) in a peer reviewed journal thus will be a relevant, though not dispositive, consideration in assessing the scientific validity of a particular technique or methodology on which an opinion is premised.

Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error, see, *e. g., United States* v. *Smith*, 869 F. 2d 348, 353–354 (CA7 1989) (surveying studies of the error rate of spectrographic voice identification technique), and the existence and maintenance of standards controlling the technique's operation, see *United States* v. *Williams*, 583 F. 2d 1194, 1198 (CA2 1978) (noting professional organization's standard governing spectrographic analysis), cert. denied, 439 U. S. 1117 (1979).

Finally, "general acceptance" can yet have a bearing on the inquiry. A "reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community." *United States* v. *Downing*, 753 F. 2d, at 1238. See also 3 Weinstein & Berger ¶ 702[03], pp. 702–41 to 702–42. Widespread acceptance can be an important factor in ruling particular evidence admissible, and "a known technique which has been able to attract only minimal support within the community," *Downing*, 753 F. 2d, at 1238, may properly be viewed with skepticism.

The inquiry envisioned by Rule 702 is, we emphasize, a flexible one.[12] Its overarching subject is the scientific valid-

[12] A number of authorities have presented variations on the reliability approach, each with its own slightly different set of factors. See, *e. g., Downing*, 753 F. 2d, at 1238–1239 (on which our discussion draws in part); 3 Weinstein & Berger ¶ 702[03], pp. 702–41 to 702–42 (on which the *Downing* court in turn partially relied); McCormick, Scientific Evidence: Defin-

ity—and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission. The focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.

Throughout, a judge assessing a proffer of expert scientific testimony under Rule 702 should also be mindful of other applicable rules. Rule 703 provides that expert opinions based on otherwise inadmissible hearsay are to be admitted only if the facts or data are "of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject." Rule 706 allows the court at its discretion to procure the assistance of an expert of its own choosing. Finally, Rule 403 permits the exclusion of relevant evidence "if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury . . . ." Judge Weinstein has explained: "Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it. Because of this risk, the judge in weighing possible prejudice against probative force under Rule 403 of the present rules exercises more control over experts than over lay witnesses." Weinstein, 138 F. R. D., at 632.

III

We conclude by briefly addressing what appear to be two underlying concerns of the parties and *amici* in this case. Respondent expresses apprehension that abandonment of "general acceptance" as the exclusive requirement for admission will result in a "free-for-all" in which befuddled juries are confounded by absurd and irrational pseudoscientific as-

---

ing a New Approach to Admissibility, 67 Iowa L. Rev. 879, 911–912 (1982); and Symposium on Science and the Rules of Evidence, 99 F. R. D. 187, 231 (1983) (statement by Margaret Berger). To the extent that they focus on the reliability of evidence as ensured by the scientific validity of its underlying principles, all these versions may well have merit, although we express no opinion regarding any of their particular details.

sertions. In this regard respondent seems to us to be overly pessimistic about the capabilities of the jury and of the adversary system generally. Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence. See *Rock* v. *Arkansas*, 483 U. S. 44, 61 (1987). Additionally, in the event the trial court concludes that the scintilla of evidence presented supporting a position is insufficient to allow a reasonable juror to conclude that the position more likely than not is true, the court remains free to direct a judgment, Fed. Rule Civ. Proc. 50(a), and likewise to grant summary judgment, Fed. Rule Civ. Proc. 56. Cf., *e. g.*, *Turpin* v. *Merrell Dow Pharmaceuticals, Inc.*, 959 F. 2d 1349 (CA6) (holding that scientific evidence that provided foundation for expert testimony, viewed in the light most favorable to plaintiffs, was not sufficient to allow a jury to find it more probable than not that defendant caused plaintiff's injury), cert. denied, 506 U. S. 826 (1992); *Brock* v. *Merrell Dow Pharmaceuticals, Inc.*, 874 F. 2d 307 (CA5 1989) (reversing judgment entered on jury verdict for plaintiffs because evidence regarding causation was insufficient), modified, 884 F. 2d 166 (CA5 1989), cert. denied, 494 U. S. 1046 (1990); Green 680–681. These conventional devices, rather than wholesale exclusion under an uncompromising "general acceptance" test, are the appropriate safeguards where the basis of scientific testimony meets the standards of Rule 702.

Petitioners and, to a greater extent, their *amici* exhibit a different concern. They suggest that recognition of a screening role for the judge that allows for the exclusion of "invalid" evidence will sanction a stifling and repressive scientific orthodoxy and will be inimical to the search for truth. See, *e. g.*, Brief for Ronald Bayer et al. as *Amici Curiae.* It is true that open debate is an essential part of both legal and scientific analyses. Yet there are important differences between the quest for truth in the courtroom and the quest for truth in the laboratory. Scientific conclusions are subject to perpetual revision. Law, on the other hand, must resolve disputes finally and quickly. The scientific project is advanced by broad and wide-ranging consideration of a multitude of hypotheses, for those that are incorrect will eventually be shown to be so, and that in itself is an advance. Conjectures that are probably wrong are of little use, however, in the project of reaching a quick, final, and binding legal judgment—often of great consequence—about a particular set of events in the past. We recognize that, in practice, a gatekeeping role for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations. That, nevertheless, is the balance that is struck by Rules of Evidence designed not for the exhaustive search for cosmic understanding but for the particularized resolution of legal disputes.[13]

IV

To summarize: "General acceptance" is not a necessary precondition to the admissibility of scientific evidence under the Federal Rules of Evidence, but the Rules of Evidence—especially Rule 702—do assign to the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand. Pertinent evidence based on scientifically valid principles will satisfy those demands.

The inquiries of the District Court and the Court of Appeals focused almost exclusively on "general acceptance," as gauged by publication and the decisions of other courts. Ac-

---

[13] This is not to say that judicial interpretation, as opposed to adjudicative factfinding, does not share basic characteristics of the scientific endeavor: "The work of a judge is in one sense enduring and in another ephemeral. . . . In the endless process of testing and retesting, there is a constant rejection of the dross and a constant retention of whatever is pure and sound and fine." B. Cardozo, The Nature of the Judicial Process 178–179 (1921).

cordingly, the judgment of the Court of Appeals is vacated, and the case is remanded for further proceedings consistent with this opinion.

It is so ordered.

Chief Justice Rehnquist, with whom Justice Stevens joins, concurring in part and dissenting in part.

The petition for certiorari in this case presents two questions: first, whether the rule of *Frye* v. *United States*, 54 App. D. C. 46, 293 F. 1013 (1923), remains good law after the enactment of the Federal Rules of Evidence; and second, if *Frye* remains valid, whether it requires expert scientific testimony to have been subjected to a peer review process in order to be admissible. The Court concludes, correctly in my view, that the *Frye* rule did not survive the enactment of the Federal Rules of Evidence, and I therefore join Parts I and II–A of its opinion. The second question presented in the petition for certiorari necessarily is mooted by this holding, but the Court nonetheless proceeds to construe Rules 702 and 703 very much in the abstract, and then offers some "general observations." *Ante,* at 593.

"General observations" by this Court customarily carry great weight with lower federal courts, but the ones offered here suffer from the flaw common to most such observations—they are not applied to deciding whether particular testimony was or was not admissible, and therefore they tend to be not only general, but vague and abstract. This is particularly unfortunate in a case such as this, where the ultimate legal question depends on an appreciation of one or more bodies of knowledge not judicially noticeable, and subject to different interpretations in the briefs of the parties and their *amici.* Twenty-two *amicus* briefs have been filed in the case, and indeed the Court's opinion contains no fewer than 37 citations to *amicus* briefs and other secondary sources.

The various briefs filed in this case are markedly different from typical briefs, in that large parts of them do not deal with decided cases or statutory language—the sort of material we customarily interpret. Instead, they deal with definitions of scientific knowledge, scientific method, scientific validity, and peer review—in short, matters far afield from the expertise of judges. This is not to say that such materials are not useful or even necessary in deciding how Rule 702 should be applied; but it is to say that the unusual subject matter should cause us to proceed with great caution in deciding more than we have to, because our reach can so easily exceed our grasp.

But even if it were desirable to make "general observations" not necessary to decide the questions presented, I cannot subscribe to some of the observations made by the Court. In Part II–B, the Court concludes that reliability and relevancy are the touchstones of the admissibility of expert testimony. *Ante,* at 590–592. Federal Rule of Evidence 402 provides, as the Court points out, that "[e]vidence which is not relevant is not admissible." But there is no similar reference in the Rule to "reliability." The Court constructs its argument by parsing the language "[i]f scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, . . . an expert . . . may testify thereto . . . ." Fed. Rule Evid. 702. It stresses that the subject of the expert's testimony must be "scientific . . . knowledge," and points out that "scientific" "implies a grounding in the methods and procedures of science" and that the word "knowledge" "connotes more than subjective belief or unsupported speculation." *Ante,* at 590. From this it concludes that "scientific knowledge" must be "derived by the scientific method." *Ibid.* Proposed testimony, we are told, must be supported by "appropriate validation." *Ibid.* Indeed, in footnote 9, the Court decides that "[i]n a case involving scientific evidence, *eviden-*

tiary reliability will be based upon *scientific validity.*" *Ante*, at 591, n. 9 (emphasis in original).

Questions arise simply from reading this part of the Court's opinion, and countless more questions will surely arise when hundreds of district judges try to apply its teaching to particular offers of expert testimony. Does all of this *dicta* apply to an expert seeking to testify on the basis of "technical or other specialized knowledge"—the other types of expert knowledge to which Rule 702 applies—or are the "general observations" limited only to "scientific knowledge"? What is the difference between scientific knowledge and technical knowledge; does Rule 702 actually contemplate that the phrase "scientific, technical, or other specialized knowledge" be broken down into numerous subspecies of expertise, or did its authors simply pick general descriptive language covering the sort of expert testimony which courts have customarily received? The Court speaks of its confidence that federal judges can make a "preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue." *Ante*, at 592–593. The Court then states that a "key question" to be answered in deciding whether something is "scientific knowledge" "will be whether it can be (and has been) tested." *Ante*, at 593. Following this sentence are three quotations from treatises, which not only speak of empirical testing, but one of which states that the " 'criterion of the scientific status of a theory is its falsifiability, or refutability, or testability.' " *Ibid.*

I defer to no one in my confidence in federal judges; but I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its "falsifiability," and I suspect some of them will be, too.

I do not doubt that Rule 702 confides to the judge some gatekeeping responsibility in deciding questions of the admissibility of proffered expert testimony. But I do not think it imposes on them either the obligation or the authority to become amateur scientists in order to perform that role. I think the Court would be far better advised in this case to decide only the questions presented, and to leave the further development of this important area of the law to future cases.

Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137, 119 S. Ct. 1167 (1999). Document in public domain.

SUPREME COURT OF THE UNITED STATES

No. 97–1709

KUMHO TIRE COMPANY, LTD., ET AL., PETITIONERS
v. PATRICK CARMICHAEL, ETC., ET AL.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE ELEVENTH CIRCUIT

[March 23, 1999]

JUSTICE BREYER delivered the opinion of the Court.

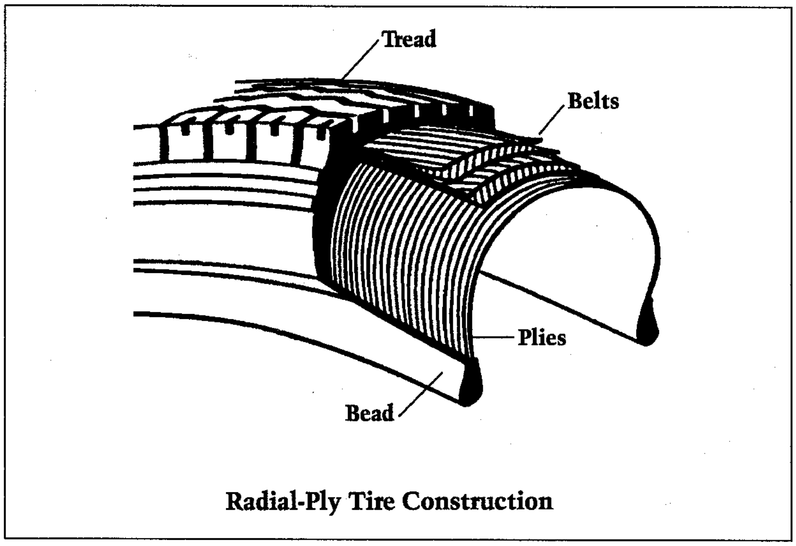
In *Daubert* v. *Merrell Dow Pharmaceuticals, Inc.,* 509 U. S. 579 (1993), this Court focused upon the admissibility of scientific expert testimony. It pointed out that such testimony is admissible only if it is both relevant and reliable. And it held that the Federal Rules of Evidence "assign to the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand." *Id.*, at 597. The Court also discussed certain more specific factors, such as testing, peer review, error rates, and "acceptability" in the relevant scientific community, some or all of which might prove helpful in determining the reliability of a particular scientific "theory or technique." *Id.,* at 593–594.

This case requires us to decide how *Daubert* applies to the testimony of engineers and other experts who are not scientists. We conclude that *Daubert*'s general holding— setting forth the trial judge's general "gatekeeping" obligation— applies not only to testimony based on "scientific" knowledge, but also to testimony based on "technical" and "other specialized" knowledge. See Fed. Rule Evid. 702. We also conclude that a trial court *may* consider one or more of the more specific factors that *Daubert* mentioned when doing so will help determine that testimony's reliability. But, as the Court stated in *Daubert*, the test of reliability is "flexible," and *Daubert*'s list of specific factors neither necessarily nor exclusively applies to all experts or in every case. Rather, the law grants a district court the same broad latitude when it decides *how* to determine reliability as it enjoys in respect to its ultimate reliability determination. See *General Electric Co.* v. *Joiner*, 522 U. S. 136, 143 (1997) (courts of appeals are to apply "abuse of discretion" standard when reviewing district court's reliability determination). Applying these standards, we determine that the District Court's decision in this case—not to admit certain expert testimony—was within its discretion and therefore lawful.

I

On July 6, 1993, the right rear tire of a minivan driven by Patrick Carmichael blew out. In the accident that followed, one of the passengers died, and others were severely injured. In October 1993, the Carmichaels brought this diversity suit against the tire's maker and its distributor, whom we refer to collectively as Kumho Tire, claiming that the tire was defective. The plaintiffs rested their case in significant part upon deposition testimony provided by an expert in tire failure analysis, Dennis Carlson, Jr., who intended to testify in support of their conclusion.

Carlson's depositions relied upon certain features of tire technology that are not in dispute. A steel-belted radial tire like the Carmichaels' is made up of a "carcass" containing many layers of flexible cords, called "plies," along which (between the cords and the outer tread) are laid steel strips called "belts." Steel wire loops, called "beads," hold the cords together at the plies' bottom edges. An outer layer, called the "tread," encases the carcass, and the entire tire is bound together in rubber, through the application of heat and various chemicals. See generally, *e.g.*, J. Dixon, Tires, Suspension and Handling 68–72 (2d ed. 1996). The bead of the tire sits upon a "bead seat," which is part of the wheel assembly. That assembly contains a "rim flange," which extends over the bead and rests against the side of the tire. See M. Mavrigian, Performance Wheels & Tires 81, 83 (1998) (illustrations).



A. Markovich, How To Buy and Care For Tires 4 (1994).

Carlson's testimony also accepted certain background facts about the tire in question. He assumed that before the blowout the tire had traveled far. (The tire was made in 1988 and had been installed some time before the Carmichaels bought the used minivan in March 1993; the Carmichaels had driven the van approximately 7,000 additional miles in the two months they had owned it.) Carlson noted that the tire's tread depth, which was 11/32 of an inch when new, App. 242, had been worn down to depths that ranged from 3/32 of an inch along some parts of the tire, to nothing at all along others. *Id.*, at 287. He conceded that the tire tread had at least two punctures which had been inadequately repaired. *Id.*, at 258–261, 322.

Despite the tire's age and history, Carlson concluded that a defect in its manufacture or design caused the blowout. He rested this conclusion in part upon three premises which, for present purposes, we must assume are not in dispute: First, a tire's carcass should stay bound to the inner side of the tread for a significant period of time after its tread depth has worn away. *Id.*, at 208–209. Second, the tread of the tire at issue had separated from its inner steel-belted carcass prior to the accident. *Id.*, at 336. Third, this "separation" caused the blowout. *Ibid.*

Carlson's conclusion that a defect caused the separation, however, rested upon certain other propositions, several of which the defendants strongly dispute. First, Carlson said that if a separation is *not* caused by a certain kind of tire misuse called "overdeflection" (which consists of underinflating the tire or causing it to carry too much weight, thereby generating heat that can undo the chemical tread/carcass bond), then, ordinarily, its cause is a tire defect. *Id.*, at 193–195, 277–278. Second, he said that if a tire has been subject to sufficient overdeflection to cause a separation, it should reveal certain physical symptoms. These symptoms include (a) tread wear on the tire's shoulder that is greater than the tread wear along the tire's center, *id.*, at 211; (b) signs of a "bead groove," where the beads have been pushed too hard against the bead seat on the inside of the tire's rim, *id.*, at 196–197; (c) sidewalls of the tire with physical signs of deterioration, such as discoloration, *id.*, at 212; and/or (d) marks on the tire's rim flange, *id.*, at 219–220. Third, Carlson said that where he does not find *at least two* of the four physical signs just mentioned (and presumably where there is no reason to suspect a less common cause of separation), he concludes that a manufacturing or design defect caused the separation. *Id.*, at 223–224.

Carlson added that he had inspected the tire in question. He conceded that the tire to a limited degree showed greater wear on the shoulder than in the center, some signs of "bead groove," some discoloration, a few marks on the rim flange, and inadequately filled puncture holes (which can also cause heat that might lead to separation). *Id.*, at 256–257, 258–261, 277, 303–304, 308. But, in each instance, he testified that the symptoms were not significant, and he explained why he believed that they did not reveal overdeflection. For example, the extra shoulder wear, he said, appeared primarily on one shoulder, whereas an overdeflected tire would reveal equally abnormal wear on both shoulders. *Id.*, at 277. Carlson concluded that the tire did not bear at least two of the four overdeflection symptoms, nor was there any less obvious cause of separation; and since neither overdeflection nor the punctures caused the blowout, a defect must have done so.

Kumho Tire moved the District Court to exclude Carlson's testimony on the ground that his methodology failed Rule 702's reliability requirement. The court agreed with Kumho that it should act as a *Daubert*-type reliability "gatekeeper," even though one might consider Carlson's testimony as "technical," rather than "scientific." See *Carmichael* v. *Samyang Tires, Inc.*, 923 F. Supp. 1514, 1521–1522 (SD Ala. 1996). The court then examined Carlson's methodology in light of the reliability-related factors that *Daubert* mentioned, such as a theory's testability, whether it "has been a subject of peer review or publication," the "known or potential rate of error," and the "degree of acceptance . . . within the relevant scientific community." 923 F. Supp., at 1520 (citing *Daubert*, 509 U. S., at 592–594). The District Court found that all those factors argued against the reliability of Carlson's methods, and it granted the motion to exclude the testimony (as well as the defendants' accompanying motion for summary judgment).

The plaintiffs, arguing that the court's application of the *Daubert* factors was too "inflexible," asked for reconsideration. And the Court granted that motion. *Carmichael* v. *Samyang Tires, Inc.*, Civ. Action No. 93–0860–CB–S (SD Ala., June 5, 1996), App. to Pet. for Cert. 1c. After reconsidering the matter, the court agreed with the plaintiffs that *Daubert* should be applied flexibly, that its four factors were simply illustrative, and that other factors could argue in favor of admissibility. It conceded that there may be widespread acceptance of a "visual-inspection method" for some relevant purposes. But the court found insufficient indications of the reliability of

> "the component of Carlson's tire failure analysis which most concerned the Court, namely, the methodology employed by the expert in analyzing the data obtained in the visual inspection, and the scientific basis, if any, for such an analysis." *Id.*, at 6c.

It consequently affirmed its earlier order declaring Carlson's testimony inadmissable and granting the defendants' motion for summary judgment.

The Eleventh Circuit reversed. See *Carmichael* v. *Samyang Tire, Inc.*, 131 F. 3d 1433 (1997). It "review[ed] . . . *de novo*" the "district court's legal decision to apply *Daubert*." *Id.*, at 1435. It noted that "the Supreme Court in *Daubert* explicitly limited its holding to cover only the 'scientific context,'" adding that "a *Daubert* analysis" applies only where an expert relies "on the application of scientific principles," rather than "on skill- or experience-based observation." *Id.*, at 1435–1436. It concluded that Carlson's testimony, which it viewed as relying on experience, "falls outside the scope of *Daubert*," that "the district court erred as a matter of law by applying *Daubert* in this case," and that the case must be remanded for further (non-*Daubert*-type) consideration under Rule 702. *Id.*, at 1436.

Kumho Tire petitioned for certiorari, asking us to determine whether a trial court "may" consider *Daubert*'s specific "factors" when determining the "admissibility of an engineering expert's testimony." Pet. for Cert. i. We granted certiorari in light of uncertainty among the lower courts about whether, or how, *Daubert* applies to expert testimony that might be characterized as based not upon "scientific" knowledge, but rather upon "technical" or "other specialized" knowledge. Fed. Rule Evid. 702; compare, *e.g.*, *Watkins* v. *Telsmith, Inc.*, 121 F. 3d 984, 990–991 (CA5 1997), with, *e.g.*, *Compton* v. *Subaru of America, Inc.*, 82 F. 3d 1513, 1518–1519 (CA10), cert. denied, 519 U. S. 1042 (1996).

II

A

In *Daubert*, this Court held that Federal Rule of Evidence 702 imposes a special obligation upon a trial judge to "ensure that any and all scientific testimony . . . is not only relevant, but reliable." 509 U. S., at 589. The initial question before us is whether this basic gatekeeping obligation applies only to "scientific" testimony or to all expert testimony. We, like the parties, believe that it applies to all expert testimony. See Brief for Petitioners 19; Brief for Respondents 17.

For one thing, Rule 702 itself says:

> "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise."

This language makes no relevant distinction between "scientific" knowledge and "technical" or "other specialized" knowledge. It makes clear that any such knowledge might become the subject of expert testimony. In *Daubert*, the Court specified that it is the Rule's word "knowledge," not the words (like "scientific") that modify that word, that "establishes a standard of evidentiary reliability." 509 U. S., at 589–590. Hence, as a matter of language, the Rule applies its reliability standard to all "scientific," "technical," or "other specialized" matters within its scope. We concede that the Court in *Daubert* referred only to "scientific" knowledge. But as the Court there said, it referred to "scientific" testimony "because that [wa]s the nature of the expertise" at issue. *Id.*, at 590, n. 8.

Neither is the evidentiary rationale that underlay the Court's basic *Daubert* "gatekeeping" determination limited to "scientific" knowledge. *Daubert* pointed out that Federal Rules 702 and 703 grant expert witnesses testimonial latitude unavailable to other witnesses on the "assumption that the expert's opinion will have a reliable basis in the knowledge and experience of his discipline." *Id.*, at 592 (pointing out that experts may testify to opinions, including those that are not based on firsthand knowledge or observation). The Rules grant that latitude to all experts, not just to "scientific" ones.

Finally, it would prove difficult, if not impossible, for judges to administer evidentiary rules under which a gatekeeping obligation depended upon a distinction between "scientific" knowledge and "technical" or "other specialized" knowledge. There is no clear line that divides the one from the others. Disciplines such as engineering rest upon scientific knowledge. Pure scientific theory itself may depend for its development upon observation and properly engineered machinery. And conceptual efforts to distinguish the two are unlikely to produce clear legal lines capable of application in particular cases. Cf.

Brief for National Academy of Engineering as *Amicus Curiae* 9 (scientist seeks to understand nature while the engineer seeks nature's modification); Brief for Rubber Manufacturers Association as *Amicus Curiae* 14–16 (engineering, as an "applied science," relies on "scientific reasoning and methodology"); Brief for John Allen et al. as *Amici Curiae* 6 (engineering relies upon "scientific knowledge and methods").

Neither is there a convincing need to make such distinctions. Experts of all kinds tie observations to conclusions through the use of what Judge Learned Hand called "general truths derived from . . . specialized experience." Hand, Historical and Practical Considerations Regarding Expert Testimony, 15 Harv. L. Rev. 40, 54 (1901). And whether the specific expert testimony focuses upon specialized observations, the specialized translation of those observations into theory, a specialized theory itself, or the application of such a theory in a particular case, the expert's testimony often will rest "upon an experience confessedly foreign in kind to [the jury's] own." *Ibid.* The trial judge's effort to assure that the specialized testimony is reliable and relevant can help the jury evaluate that foreign experience, whether the testimony reflects scientific, technical, or other specialized knowledge.

We conclude that *Daubert*'s general principles apply to the expert matters described in Rule 702. The Rule, in respect to all such matters, "establishes a standard of evidentiary reliability." 509 U. S., at 590. It "requires a valid . . . connection to the pertinent inquiry as a precondition to admissibility." *Id.*, at 592. And where such testimony's factual basis, data, principles, methods, or their application are called sufficiently into question, see Part III, *infra*, the trial judge must determine whether the testimony has "a reliable basis in the knowledge and experience of [the relevant] discipline." 509 U. S., at 592.

B

The petitioners ask more specifically whether a trial judge determining the "admissibility of an engineering expert's testimony" *may* consider several more specific factors that *Daubert* said might "bear on" a judge's gatekeeping determination. These factors include:

– Whether a "theory or technique . . . can be (and has been) tested";

– Whether it "has been subjected to peer review and publication";

– Whether, in respect to a particular technique, there is a high "known or potential rate of error" and whether there are "standards controlling the technique's operation"; and

– Whether the theory or technique enjoys "general acceptance" within a "relevant scientific community." 509 U. S., at 592–594.

Emphasizing the word "may" in the question, we answer that question yes.

Engineering testimony rests upon scientific foundations, the reliability of which will be at issue in some cases. See, *e.g.*, Brief for Stephen Bobo et al. as *Amici Curiae* 23 (stressing the scientific bases of engineering disciplines). In other cases, the relevant reliability concerns may focus upon personal knowledge or experience. As the Solicitor General points out, there are many different kinds of experts, and many different kinds of expertise. See Brief for United States as *Amicus Curiae* 18–19, and n. 5 (citing cases involving experts in drug terms, handwriting analysis, criminal *modus operandi*, land valuation, agricultural practices, railroad procedures, attorney's fee valuation, and others). Our emphasis on the word "may" thus reflects *Daubert*'s description of the Rule 702 inquiry as "a flexible one." 509 U. S., at 594. *Daubert* makes clear that the factors it mentions do *not* constitute a "definitive checklist or test." *Id.*, at 593. And *Daubert* adds that the gatekeeping inquiry must be "'tied to the facts'" of a particular "case." *Id.*, at 591 (quoting *United States* v. *Downing*, 753 F. 2d 1224, 1242 (CA3 1985)). We agree with the Solicitor General that "[t]he factors identified in *Daubert* may or may not be pertinent in assessing reliability, depending on the nature of the issue, the expert's particular expertise, and the subject of his testimony." Brief for United States as *Amicus Curiae* 19. The conclusion, in our view, is that we can neither rule out, nor rule in, for all cases and for all time the applicability of the factors mentioned in *Daubert*, nor can we now do so for subsets of cases categorized by category of expert or by kind of evidence. Too much depends upon the particular circumstances of the particular case at issue.

Daubert itself is not to the contrary. It made clear that its list of factors was meant to be helpful, not definitive. Indeed, those factors do not all necessarily apply even in every instance in which the reliability of scientific testimony is challenged. It might not be surprising in a particular case, for example, that a claim made by a scientific witness has never been the subject of peer review, for the particular application at issue may never previously have interested any scientist. Nor, on the other hand, does the presence of *Daubert*'s general acceptance factor help show that an expert's testimony is reliable where the discipline itself lacks reliability, as, for example, do theories grounded in any so-called generally accepted principles of astrology or necromancy.

At the same time, and contrary to the Court of Appeals' view, some of *Daubert*'s questions can help to evaluate the reliability even of experience-based testimony. In certain cases, it will be appropriate for the trial judge to ask, for example, how often an engineering expert's experience-based methodology has produced erroneous results, or whether such a method is generally accepted in the relevant engineering community. Likewise, it will at times be useful to ask even of a witness whose expertise is based purely on experience, say, a perfume tester able to distinguish among 140 odors at a sniff, whether his preparation is of a kind that others in the field would recognize as acceptable.

We must therefore disagree with the Eleventh Circuit's holding that a trial judge may ask questions of the sort *Daubert* mentioned only where an expert "relies on the application of scientific principles," but not where an expert relies "on skill- or experience-based observation." 131 F. 3d, at 1435. We do not believe that Rule 702 creates a schematism that segregates expertise by type while mapping certain kinds of questions to certain kinds of experts. Life and the legal cases that it generates are too complex to warrant so definitive a match.

To say this is not to deny the importance of *Daubert*'s gatekeeping requirement. The objective of that requirement is to ensure the reliability and relevancy of expert testimony. It is to make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field. Nor do we deny that, as stated in *Daubert*, the particular questions that it mentioned will often be appropriate for use in determining the reliability of challenged expert testimony. Rather, we conclude that the trial judge must have considerable leeway in deciding in a particular case how to go about determining whether particular expert testimony is reliable. That is to say, a trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of expert testimony.

C

The trial court must have the same kind of latitude in deciding *how* to test an expert's reliability, and to decide whether or when special briefing or other proceedings are needed to investigate reliability, as it enjoys when it decides *whether* that expert's relevant testimony is reliable. Our opinion in *Joiner* makes clear that a court of appeals is to apply an abuse-of-discretion standard when it "review[s] a trial court's decision to admit or exclude expert testimony." 522 U. S., at 138–139. That standard applies as much to the trial court's decisions about how to determine reliability as to its ultimate conclusion. Otherwise, the trial judge would lack the discretionary authority needed both to avoid unnecessary "reliability" proceedings in ordinary cases where the reliability of an expert's methods is properly taken for granted, and to require appropriate proceedings in the less usual or more complex cases where cause for questioning the expert's reliability arises. Indeed, the Rules seek to avoid "unjustifiable expense and delay" as part of their search for "truth" and the "jus[t] determin[ation]" of proceedings. Fed. Rule Evid. 102. Thus, whether *Daubert*'s specific factors are, or are not, reasonable measures of reliability in a particular case is a matter that the law grants the trial judge broad latitude to determine. See *Joiner*, *supra*, at 143. And the Eleventh Circuit erred insofar as it held to the contrary.

III

We further explain the way in which a trial judge "may" consider *Daubert*'s factors by applying these considerations to the case at hand, a matter that has been briefed exhaustively by the parties and their 19 *amici*. The District Court did not doubt Carlson's qualifications, which included a masters degree in mechanical engineering, 10 years' work at Michelin America, Inc., and testimony as a tire failure consultant in other tort cases. Rather, it excluded the testimony because, despite those qualifications, it initially doubted, and then found unreliable, "the methodology employed by the expert in analyzing the data obtained in the visual inspection, and the scientific basis, if any, for such an analysis." Civ. Action No. 93–0860–CB–S (SD Ala., June 5, 1996), App. to Pet. for Cert. 6c. After examining the transcript in "some detail," 923 F. Supp., at 1518–519, n. 4, and after considering respondents' defense of Carlson's methodology, the District Court determined that Carlson's testimony was not reliable. It fell outside the range where experts might reasonably differ, and where the jury must decide among the conflicting views of different experts, even though the evidence is "shaky." *Daubert*, 509 U. S., at 596. In our view, the doubts that triggered the District Court's initial inquiry here were reasonable, as was the court's ultimate conclusion.

For one thing, and contrary to respondents' suggestion, the specific issue before the court was not the reasonableness *in general* of a tire expert's use of a visual and tactile inspection to determine whether overdeflection had caused the tire's tread to separate from its steel-belted carcass. Rather, it was the reasonableness of using such an approach, along with Carlson's particular method of analyzing the data thereby obtained, to draw a conclusion regarding *the particular matter to which the expert testimony was directly relevant*. That matter concerned the likelihood that a defect in the tire at issue caused its tread to separate from its carcass. The tire in question, the expert conceded, had traveled far enough so that some of the tread had been worn bald; it should have been taken out of service; it had been repaired (inadequately) for punctures; and it bore some of the very marks that the expert said indicated, not a defect, but abuse through overdeflection. See *supra*, at 3–5; App. 293–294. The relevant issue was whether the expert could reliably determine the cause of *this* tire's separation.

Nor was the basis for Carlson's conclusion simply the general theory that, in the absence of evidence of abuse, a defect will normally have caused a tire's separation. Rather, the expert employed a more specific theory to establish the existence (or absence) of such abuse. Carlson testified precisely that in the absence of *at least two* of four signs of abuse (proportionately greater tread wear on the shoulder; signs of grooves caused by the beads; discolored sidewalls; marks on the rim flange) he concludes that a defect caused the separation. And his analysis depended upon acceptance of a further implicit proposition, namely, that his visual and tactile inspection could determine that the tire before him had not been abused despite some evidence of the presence of the very signs for which he looked (and two punctures).

For another thing, the transcripts of Carlson's depositions support both the trial court's initial uncertainty and its final conclusion. Those transcripts cast considerable doubt upon the reliability of both the explicit theory (about the need for two signs of abuse) and the implicit proposition (about the significance of visual inspection in this case). Among other things, the expert could not say whether the tire had traveled more than 10, or 20, or 30, or 40, or 50 thousand miles, adding that 6,000 miles was "about how far" he could "say with any certainty." *Id.*, at 265. The court could reasonably have wondered about the reliability of a method of visual and tactile inspection sufficiently precise to ascertain with some certainty the abuse-related significance of minute shoulder/center relative tread wear differences, but insufficiently precise to tell "with any certainty" from the tread wear whether a tire had traveled less than 10,000 or more than 50,000 miles. And these concerns might have been augmented by Carlson's repeated reliance on the "subjective[ness]" of his mode of analysis in response to questions seeking specific information regarding how he could differentiate between a tire that actually had been overdeflected and a tire that merely looked as though it had been. *Id.*, at 222, 224–225, 285–286. They would have been further augmented by the fact that Carlson said he had inspected the tire itself for the first time the morning of his first deposition, and then only for a few hours. (His initial conclusions were based on photographs.) *Id.*, at 180.

Moreover, prior to his first deposition, Carlson had issued a signed report in which he concluded that the tire had "not been . . . overloaded or underinflated," not because of the absence of "two of four" signs of abuse, but simply because "the rim flange impressions . . . were normal." *Id.*, at 335–336. That report also said that the "tread depth remaining was 3/32 inch," *id.*, at 336, though the opposing expert's (apparently undisputed) measurements indicate that the tread depth taken at various positions around the tire actually ranged from .5/32 of an inch to 4/32 of an inch, with the tire apparently showing greater wear along *both* shoulders than along the center, *id.*, at 432–433.

Further, in respect to one sign of abuse, bead grooving, the expert seemed to deny the sufficiency of his own simple visual-inspection methodology. He testified that most tires have some bead groove pattern, that where there is reason to suspect an abnormal bead groove he would ideally "look at a lot of [similar] tires" to know the grooving's significance, and that he had not looked at many tires similar to the one at issue. *Id.*, at 212–213, 214, 217.

Finally, the court, after looking for a defense of Carlson's methodology as applied in these circumstances, found no convincing defense. Rather, it found (1) that "none" of the *Daubert* factors, including that of "general acceptance" in the relevant expert community, indicated that Carlson's testimony was reliable, 923 F. Supp., at 1521; (2) that its own analysis "revealed no countervailing factors operating in favor of admissibility which could outweigh those identified in *Daubert*," App. to Pet. for Cert. 4c; and (3) that the "parties identified no such factors in their briefs," *ibid.* For these three reasons *taken together*, it concluded that Carlson's testimony was unreliable.

Respondents now argue to us, as they did to the District Court, that a method of tire failure analysis that employs a visual/tactile inspection is a reliable method, and they point both to its use by other experts and to Carlson's long experience working for Michelin as sufficient indication that that is so. But no one denies that an expert might draw a conclusion from a set of observations based on extensive and specialized experience. Nor does anyone deny that, as a general matter, tire abuse may often be identified by qualified experts through visual or tactile inspection of the tire. See Affidavit of H. R. Baumgardner 1–2, cited in Brief for National Academy of Forensic Engineers as *Amici Curiae* 16 (Tire engineers rely on visual examination and process of elimination to analyze experimental test tires). As we said before, *supra*, at 14, the question before the trial court was specific, not general. The trial court had to decide whether this particular expert had sufficient specialized knowledge to assist the jurors "in deciding the particular issues in the case." 4 J. McLaughlin, Weinstein's Federal Evidence ¶702.05[1], p. 702–33 (2d ed. 1998); see also Advisory Committee's Note on Proposed Fed. Rule Evid. 702, Preliminary Draft of Proposed Amendments to the Federal Rules of Civil Procedure and Evidence: Request for Comment 126 (1998) (stressing that district courts must "scrutinize" whether the "principles and methods" employed by an expert "have been properly applied to the facts of the case").

The particular issue in this case concerned the use of Carlson's two-factor test and his related use of visual/tactile inspection to draw conclusions on the basis of what seemed small observational differences. We have found no indication in the record that other experts in the industry use Carlson's two-factor test or that tire experts such as Carlson normally make the very fine distinctions about, say, the symmetry of comparatively greater shoulder tread wear that were necessary, on Carlson's own theory, to support his conclusions. Nor, despite the prevalence of tire testing, does anyone refer to any articles or papers that validate Carlson's approach. Compare Bobo, Tire Flaws and Separations, in Mechanics of Pneumatic Tires 636–637 (S. Clark ed. 1981); C. Schnuth et al., Compression Grooving and Rim Flange Abrasion as Indicators of Over-Deflected Operating Conditions in Tires, presented to Rubber Division of the American Chemical Society, Oct. 21–24, 1997; J. Walter & R. Kiminecz, Bead Contact Pressure Measurements at the Tire-Rim Interface, presented to Society of Automotive Engineers, Feb. 24–28, 1975. Indeed, no one has argued that Carlson himself, were he still working for Michelin, would have concluded in a report to his employer that a similar tire was similarly defective on grounds identical to those upon which he rested his conclusion here. Of course, Carlson himself claimed that his method was accurate, but, as we pointed out in *Joiner*, "nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert." 522 U. S., at 146.

Respondents additionally argue that the District Court too rigidly applied *Daubert*'s criteria. They read its opinion to hold that a failure to satisfy any one of those criteria automatically renders expert testimony inadmissible. The District Court's initial opinion might have been vulnerable to a form of this argument. There, the court, after rejecting respondents' claim that Carlson's testimony was "exempted from *Daubert*-style scrutiny" because it was "technical analysis" rather than "scientific evidence," simply added that "none of the four admissibility criteria outlined by the *Daubert* court are satisfied." 923 F. Supp., at 1522. Subsequently, however, the court granted respondents' motion for reconsideration. It then explicitly recognized that the relevant reliability inquiry "should be 'flexible,'" that its "'overarching subject [should be] . . . validity' and reliability," and that "*Daubert* was intended neither to be exhaustive nor to apply in every case." App. to Pet. for Cert. 4c (quoting *Daubert*, 509 U. S., at 594–595). And the court ultimately based its decision upon Carlson's failure to satisfy either *Daubert*'s factors *or any other* set of reasonable reliability criteria. In light of the record as developed by the parties, that conclusion was within the District Court's lawful discretion.

In sum, Rule 702 grants the district judge the discretionary authority, reviewable for its abuse, to determine reliability in light of the particular facts and circumstances of the particular case. The District Court did not abuse its discretionary authority in this case. Hence, the judgment of the Court of Appeals is

Reversed.

SUPREME COURT OF THE UNITED STATES

No. 97–1709

KUMHO TIRE COMPANY, LTD., ET AL., PETITIONERS *v.* PATRICK CARMICHAEL, ETC., ET AL.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE ELEVENTH CIRCUIT

[March 23, 1999]

JUSTICE SCALIA, with whom JUSTICE O'CONNOR and JUSTICE THOMAS join, concurring.

I join the opinion of the Court, which makes clear that the discretion it endorses—trial-court discretion in choosing the manner of testing expert reliability—is not discretion to abandon the gatekeeping function. I think it worth adding that it is not discretion to perform the function inadequately. Rather, it is discretion to choose among *reasonable* means of excluding expertise that is *fausse* and science that is junky. Though, as the Court makes clear today, the *Daubert* factors are not holy writ, in a particular case the failure to apply one or another of them may be unreasonable, and hence an abuse of discretion.

SUPREME COURT OF THE UNITED STATES

No. 97–1709

KUMHO TIRE COMPANY, LTD., ET AL., PETITIONERS *v.* PATRICK CARMICHAEL, ETC., ET AL.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE ELEVENTH CIRCUIT

[March 23, 1999]

JUSTICE STEVENS, concurring in part and dissenting in part.

The only question that we granted certiorari to decide is whether a trial judge "[m]ay . . . consider the four factors set out by this Court in *Daubert* v. *Merrill Dow Pharmaceuticals, Inc.*, 509 U. S. 579 (1993), in a Rule 702 analysis of admissibility of an engineering expert's testimony." Pet. for Cert. i. That question is fully and correctly answered in Parts I and II of the Court's opinion, which I join.

Part III answers the quite different question whether the trial judge abused his discretion when he excluded the testimony of Dennis Carlson. Because a proper answer to that question requires a study of the record that can be performed more efficiently by the Court of Appeals than by the nine Members of this Court, I would remand the case to the Eleventh Circuit to perform that task. There are, of course, exceptions to most rules, but I firmly believe that it is neither fair to litigants nor good practice for this Court to reach out to decide questions not raised by the certiorari petition. See *General Electric Co.* v. *Joiner*, 522 U. S. 136, 150–151 (1997) (STEVENS, J., concurring in part and dissenting in part).

Accordingly, while I do not feel qualified to disagree with the well-reasoned factual analysis in Part III of the Court's opinion, I do not join that Part, and I respectfully dissent from the Court's disposition of the case.

U.S. v. Herrera, 704 F. 3d 480 (2013). Document in public domain.

In the

United States Court of Appeals

For the Seventh Circuit

No. 11-2894

UNITED STATES OF AMERICA,

Plaintiff-Appellee,

v.

CLACY WATSON HERRERA,

Defendant-Appellant.

Appeal from the United States District Court
for the Northern District of Illinois, Eastern Division.
No. 01 CR 1098-1—**Rebecca R. Pallmeyer**, *Judge.*

ARGUED SEPTEMBER 28, 2012—DECIDED JANUARY 9, 2013

Before POSNER, ROVNER, and SYKES, *Circuit Judges*.

POSNER, *Circuit Judge*. Two years ago, in response to a petition for a writ of mandamus filed by the government during the criminal trial of the defendant on drug charges, we ordered the district court to admit into evidence an exhibit labeled "Roberson Seizure 2"; to allow the government to recall Stephen Koop to testify at trial about the recovery of latent fingerprints from that exhibit; and to allow testimony regarding comparison of the latent prints with patent fingerprints known to be the defendant's. *In re United States*, 614 F.3d 661 (7th Cir. 2010). The judge had excluded the exhibit and related testimony because he suspected, though on the most tenuous of grounds, that the government had tampered with the fingerprint evidence. He threatened to grant the defendant's request for a mistrial on the ground of prosecutorial misconduct that was (the judge believed) intended to avert a likely acquittal, a ground that if sustained would have barred any further prosecution of the defendant as placing him in double jeopardy. *Oregon v. Kennedy*, 456 U.S. 667, 679 (1982); *United States v. Catton*, 130 F.3d 805, 807-08 (7th Cir. 1997); see also *United States v. Buljubasic*, 808 F.2d 1260, 1265 (7th Cir. 1987). We also ordered the case reassigned to another district judge. This was done and the trial, which had been interrupted by the mandamus proceeding, resumed, and ended shortly in the conviction of the defendant. The judge sentenced him to 340 months in prison for a variety of drug-related offenses. He appeals.

Many of his arguments repeat ones he made in the mandamus proceeding. (In effect he is asking us to rehear our previous decision—two years after the deadline for asking for rehearing expired.) The only such argument that we didn't discuss is based on *Will v. United States*, 389 U.S. 90, 96-97 (1967), which forbids the use of mandamus as a substitute for an appeal that is forbidden—and the government is not permitted to appeal an evidentiary ruling in a criminal case once the trial has begun. 18 U.S.C. § 3731. But the Court in *Will* held only that the court of appeals hadn't explained why the district court's ordering the government to give the defendant a bill of particulars was so "seriously disruptive of the efficient administration of criminal justice in the Northern District of Illinois" as to warrant mandamus. 389 U.S. at 104. The district judge's order in the present case was no run-of-the-mill mistaken procedural or evidentiary ruling. The order seriously disrupted the prosecution's case, and did so, as we are about to show, on the basis of utterly baseless but damaging imputations of grave (criminal, really) prosecutorial misconduct; involved the flouting of governing precedents; and would probably have resulted in a groundless acquittal. The order thus warranted correction by mandamus. See *United States v. Vinyard*, 539 F.3d 589, 591-92 (7th Cir. 2008).

The chain of events that culminated in the mandamus proceeding had begun with the district judge's decision to exclude evidence that two of the defendant's fingerprints had been recovered from a bag of heroin wrapped in tape and further encased in condoms and found in a drug courier's rectum. The heroin had been removed from the bag and placed in an evidence bag and then both it and the packaging (the tape and condoms) had been placed in another evidence bag and it was this second exhibit that was at issue. The district judge's ground for excluding it was his belief that the government hadn't adequately demonstrated the requisite "chain of custody"—hadn't demonstrated that there had been no opportunity to tamper with or otherwise mishandle the evidence between the time it was obtained and the trial. The judge made this ruling in the face of the government's having offered ten witnesses to establish that the chain of custody had remained intact.

The judge was disturbed because the exhibit had, according to an evidence log sheet, gained 20 grams in weight between May and September 2001. (Yet he attached no significance to its having gained 190 grams between September 2001 and the trial.) He thought the weight gain might have been attributable to federal officers' pressing a piece of adhesive tape containing the defendant's fingerprints (obtained elsewhere) onto the packaging of the heroin. That suspicion grew into a conviction, for which there was no rational basis, that government lawyers had lied about the chain of custody. To no avail the government explained that the reason for the increase in weight was that the bag with the fingerprints, after being opened so that the presence and amount of the illegal drug contained in it could be verified, and later closed up again, had been weighed together with other bags. The reported weight was the weight of the package containing all the bags, and thus there were more bags in it. Obviously the package would not have gained 210 grams (20 + 190)—almost half a pound—from replacing a piece of the tape in which one of the bags was wrapped by a piece of tape containing the defendant's fingerprints.

The judge acknowledged that his supposition of tampering was "speculative." That was an understatement. For among other things the defendant had not been extradited to the United States until long after the alleged tampering, and until he was extradited the government did not have a set of fingerprints known to be his. And no one has explained how fingerprints on another piece of material could have been transferred to the *adhesive* side of the tape, which was where they were found. It's one thing to press your finger on the adhesive side of a tape and remove the finger, leaving a print, but another thing to press a piece of paper containing your fingerprint on the adhesive side of the tape—try removing the paper without destroying the print.

The defendant's petition, and amended petition, for rehearing did not defend the judge's conjecture that the weight discrepancy indicated tampering. We concluded that while the defendant could argue at trial that the jury should disregard the fingerprint evidence, there was no justification for excluding it in advance of trial on the "speculative" ground excogitated by the judge. Once the government presents evidence, as it did here (remember the ten witnesses), that adequate precautions had been taken to preserve the evidence challenged by the defendant, it has established admissibility, though at trial the defendant can challenge the adequacy of the precautions and present evidence of tampering. *United States v. Lee*, 502 F.3d 691, 697-98 (7th Cir. 2007); *United States v. Kelly*, 14 F.3d 1169, 1175 (7th Cir. 1994); *United States v. Brumfield*, 686 F.3d 960, 965 (8th Cir. 2012); see also *Melendez-Diaz v. Massachusetts*, 557 U.S. 305, 311 n. 1 (2009). And that means by the way that even if our mandamus order was ultra vires it didn't undermine the fairness of the trial or the justice of the defendant's conviction. The fingerprint evidence should not have been excluded, and once admitted confirmed his guilt. We take up at the end of our opinion the defendant's distinct argument that the reassignment of the case to another judge prejudiced the jury, and show that that argument has no merit either.

The fresh issue relating to the fingerprint evidence is whether the prints of two fingers found on the adhesive tape were the defendant's. They were latent rather than patent fingerprints. Patent fingerprints are made by pressing a fingertip covered with ink on a white card or similar white surface, and are visible. Latent fingerprints are prints, usually invisible, left on a smooth surface when a person touches it with a finger or fingers. Laboratory techniques are employed to make a latent fingerprint visible so that it can be compared with other fingerprints. The latent prints on the adhesive tape on the bag of heroin in this case were found by a fingerprint examiner to match the defendant's patent prints made in the course of the criminal investigation, and the government therefore offered the match as evidence of the defendant's participation in the drug ring. The defendant argues that methods of matching latent prints with other latent prints or with patent prints have not been shown to be reliable enough to be admissible as evidence under the standard for reliability set forth in Fed. R. Evid. 702, 703; *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 592-93 (1993); and *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 149 (1999).

The method the examiner used is called ACE-V and is the standard method for determining whether two fingerprints are from the same person. See Scientific Working Group on Friction Ridge Analysis, Study and Technology, "Standards for Examining Friction Ridge Impressions and Resulting Conclusions," Sept. 13, 2011, www.swgfast.org/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf (visited Jan. 4, 2013); Michele Triplett & Lauren Cooney, "The Etiology of ACE-V and Its Proper Use: An Exploration of the Relationship Between ACE-V and the Scientific Method of Hypothesis Testing," 56 *J. Forensic Identification* 345, 346 (2006). The defendant is therefore mounting a frontal assault on the use of fingerprint evidence in litigation, an attack the courts have frequently rebuffed. See, e.g., *United States v. Havvard*, 260 F.3d 597, 601 (7th Cir. 2001); *United States v. George*, 363 F.3d 666, 672-73 (7th Cir. 2004); *United States v. Crisp*, 324 F.3d 261, 268-70 (4th Cir. 2003); *United States v. Mitchell*, 365 F.3d 215, 235-46 (3d Cir. 2004).

ACE-V is an acronym for analysis, comparison, evaluation, and verification, and has been described as follows:

> The process begins with the analysis of the unknown friction ridge print (now often a digital image of a latent print). Many factors affect the quality and quantity of detail in the latent print and also introduce variability in the resulting impression . . . . If the examiner deems that there is sufficient detail in the latent print (and the known prints), the comparison of the latent print to the known prints begins.
>
> Visual comparison consists of discerning, visually "measuring," and comparing—within the comparable areas of the latent print and the known prints—the

> details that correspond. The amount of friction ridge detail available for this step depends on the clarity of the two impressions. The details observed might include the overall shape of the latent print, anatomical aspects, ridge flows, ridge counts, shape of the core, delta location and shape, lengths of the ridges, minutia location and type, thickness of the ridges and furrows, shapes of the ridges, pore position, crease patterns and shapes, scar shapes, and temporary feature shapes (e.g., a wart).
>
> At the completion of the comparison, the examiner performs an evaluation of the agreement of the friction ridge formations in the two prints and evaluates the sufficiency of the detail present to establish an identification (source determination). Source determination is made when the examiner concludes, based on his or her experience, that sufficient quantity and quality of friction ridge detail is in agreement between the latent print and the known print. Source exclusion is made when the process indicates sufficient disagreement between the latent print and known print. If neither an identification nor an exclusion can be reached, the result of the comparison is inconclusive. Verification occurs when another qualified examiner repeats the observations and comes to the same conclusion, although the second examiner may be aware of the conclusion of the first.

National Research Council of the National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward* 137-38 (2009).

The methodology requires recognizing and categorizing scores of distinctive features in the prints, see Davide Maltoni et al., *Handbook of Fingerprint Recognition* 97-101 (2d ed. 2009); Federal Bureau of Investigation, *The Science of Fingerprints: Classification and Uses* 5-86 (2006), and it is the distinctiveness of these features, rather than the ACE-V method itself, that enables expert fingerprint examiners to match fingerprints with a high degree of confidence. That's not to say that fingerprint matching (especially when it involves latent fingerprints, as in this case) is as reliable as DNA evidence, for example. Forensic DNA analysis involves comparing a strand of DNA (the genetic code) from the suspect with a strand of DNA found at the crime scene. The comparison is done with scientific instruments and determines whether the segments are chemically identical. Errors are vanishingly rare provided that the strands of code are reasonably intact. As we explained in *United States v. Ford*, 683 F.3d 761, 768 (7th Cir. 2012),

> What is involved, very simply, in forensic DNA analysis is comparing a strand of DNA (the genetic code) from the suspect with a strand of DNA found at the crime scene. See "DNA Profiling," *Wikipedia*, http://en.wikipedia.org/wiki/DNA_profiling (visited May 31, 2012). Comparisons are made at various locations on each strand. At each location there is an allele (a unique gene form). In one location, for example, the probability of a person's having a particular allele might be 7 percent, and in another 10 percent. Suppose that the suspect's DNA and the DNA at the crime scene contained the same alleles

> at each of the two locations. The probability that the DNA was someone else's would be 7 percent if the comparison were confined to the first location, but only .7 percent (7 percent of 10 percent) if the comparison were expanded to two locations, because the probabilities are independent. Suppose identical alleles were found at 10 locations, which is what happened in this case; the probability that two persons would have so many identical alleles, a probability that can be computed by multiplying together the probabilities of an identical allele at each location, becomes infinitesimally small—in fact 1 in 29 trillion, provided no other comparisons reveal that the alleles at the same location on the two strands of DNA are different. This is the same procedure used for determining the probability that a perfectly balanced coin flipped 10 times in a row will come up heads all 10 times. The probability is $.5^{10}$, which is less than 1 in 1000.

Chemical tests can determine whether two alleles are identical, but a fingerprint analyst must visually recognize and classify the relevant details in the latent print—which is difficult if the print is incomplete or smudged. "[T]he assessment of latent prints from crime scenes is based largely on human interpretation. . . . [T]he process does not allow one to stipulate specific measurements in advance, as is done for a DNA analysis. Moreover, a small stretching of distance between two fingerprint features, or a twisting of angles, can result from either a difference between the fingers that left the prints or from distortions from the impression process." National Research Council, *supra*, at 139.

Matching latent fingerprints is thus a bit like an opinion offered by an art expert asked whether an unsigned painting was painted by the known painter of another painting; he makes or rejects a match on the basis of visual evidence. Eyewitness evidence is similar. The eyewitness saw the perpetrator of a crime. His recollection of the perpetrator's appearance is analogous to a latent fingerprint. He sees the defendant at the trial—that sighting is analogous to a patent fingerprint. He is asked to match his recollection against the courtroom sighting—and he is allowed to testify that the defendant *is* the perpetrator, not just that there is a close resemblance. A lineup, whether photo or in-person, is a related method of adducing matching evidence, as is handwriting evidence.

Matching evidence of the kinds that we've just described, including fingerprint evidence, is less rigorous than the kind of scientific matching involved in DNA evidence; eyewitness evidence is not scientific at all. But no one thinks that only scientific evidence may be used to convict or acquit a defendant. The increasingly well documented fallibility of eyewitness testimony, see Elizabeth F. Loftus *et al.*, *Eyewitness Testimony: Civil and Criminal* (4th ed. 2007); *United States v. Ford*, *supra*, 683 F.3d at 764-66, has not banished it from criminal trials. *Perry v. New Hampshire*, 132 S. Ct. 716, 728 (2012).

Evidence doesn't have to be infallible to be probative. Probability of guilt is a function of all the evidence in a case, and if items of evidence are independent of one another in the sense that the truth of any one item is not influenced by the truth of any other, the probability of guilt may be much higher if there is evidence from many independent sources (several eyewitnesses, an eyewitness plus fingerprints, etc.) than it would be were there only the evidence of one eyewitness, say. If "the prosecution submits three items of evidence of the defendant's guilt (and the defendant submits no evidence of his innocence), and the probability that item 1 is spurious is 10 percent, the probability that item 2 is spurious is also 10 percent, and likewise item 3 [, then the] probability that all three are spurious (assuming that the probabilities are independent—that is, that the probability that one piece of evidence is spurious does not affect the probability that another is), and therefore that the defendant should be acquitted, is only one in a thousand (.1 x .1 x .1)." *United States v. Williams*, 698 F.3d 374, 379 (7th Cir. 2012).

The defendant intimates that any evidence that requires the sponsorship of an expert witness, as fingerprint evidence does, must be found to be good science before it can be admitted under the doctrine of the *Daubert* case and Rules 702 or 703 of the Federal Rules of Evidence. But expert evidence is not limited to "scientific" evidence, however such evidence might be defined. *Kumho Tire Co. v. Carmichael*, *supra*, 526 U.S. at 150-51; *Tuf Racing Products, Inc. v. American Suzuki Motor Corp.*, 223 F.3d 585, 591 (7th Cir. 2000). It includes any evidence created or validated by expert methods and presented by an expert witness that is shown to be reliable. In a case involving an alleged forgery of a painting, there might be expert scientific evidence based on tests of the age of the canvas or paint; but there might also be expert evidence, offered by a dealer or art historian or other art expert, on the style of a particular artist. That evidence would be the expert's opinion, based on comparison with other paintings, of the genuineness of the painting alleged to be a forgery. See, e.g., *Levin v. Dalva Brothers, Inc.*, 459 F.3d 68, 78-79 (1st Cir. 2006); *United States v. Tobin*, 576 F.2d 687, 690-91, 693 (5th Cir. 1978).

Fingerprint experts such as the government's witness in this case—who has been certified as a latent print examiner by the International Association for Identification, the foremost international fingerprint organization (there are only about 840 IAI-certified latent examiners in the world, out of 15,000 total examiners)—receive extensive training; and errors in fingerprint matching by expert examiners appear to be very rare. Of the first 194 prisoners in the United States exonerated by DNA evidence, none had been convicted on the basis of erroneous fingerprint matches, whereas 75 percent had been convicted on the basis of mistaken eyewitness identification. Greg Hampikian *et al.*, "The Genetics of Innocence: Analysis of 194 U.S. DNA Exonerations," 12 *Annual Rev. of Genomics and Human Genetics* 97, 106 (2011). The probability of two people in the world having identical fingerprints is not known, but it appears to be extremely low. Steven M. Stigler, "Galton and Identification by Fingerprints," 140 *Genetics* 857, 858 (1995); David A. Stoney & John I. Thornton, "A Critical Analysis of Quantitative Fingerprint Individuality Models," 31 *J. of Forensic Sciences* 1187 (1986). The great statistician Francis Galton estimated the probability as 1 in 64 billion. Galton, *Finger Prints* 110 (1892); Stigler, *supra* at 858. That was not an estimate of the probability of a mistaken matching of a latent to a patent or another latent fingerprint. Yet errors in such matching appear to be very rare, though the matching process is judgmental rather than scientifically rigorous because it depends on how readable the latent fingerprint is and also on how distorted a version of the person's patent fingerprint it is. Examiners' training includes instruction on how to determine whether a latent print contains enough detail to enable a reliable matching to another print. Ultimately the matching depends on "subjective judgments by the examiner," National Research Council, *supra*, at 139, but responsible fingerprint matching is admissible evidence, in general and in this case.

The other issues presented by the appeal that merit discussion arise from the interruption of the trial by the mandamus proceeding and the resulting reassignment of the case to a different district judge. The consequence was an eleven-day hiatus in the trial. The defendant argues that when the trial resumed, the jurors, remembering the skeptical remarks that the original judge had made about the government's evidence, must have thought that he had been "punished" for siding with the defendant by being removed and therefore that the jury should convict. That is unpersuasive conjecture. Because of sickness most commonly, but sometimes for other reasons, such as belated discovery of a ground for recusal, a judge is sometimes replaced during a trial and when that happens the new judge tells the jury that such replacements happen occasionally and the jurors are not to worry about the change in judges or speculate about the reason for it. The new judge in this case didn't explain the cause of the delays but did say:

> It is very important for me to emphasize this instruction, that however you may feel about the delays in this case, you are not to hold those feelings against anybody in this courtroom . . . . In fact, I am going to instruct you right now that you not speculate about the causes or reasons for the delays at all . . . . To the extent that you have been told or you have come to believe that the delays are somehow the fault of the government or the fault of the defense counsel, I am instructing you that you put those concerns out of your mind completely . . . . At the end of this case, we will not be asking you, did the trial go smoothly? And if not, whose fault was it? That will not be a question you will be asked to consider. The only question you will be asked to consider at the conclusion of this case is, did the government meet its burden of proof? That's the only question. And concerns about delays are not to be in your mind at all . . . . From time to time there are reasons that we have to interrupt the smooth progress of a trial. It's happened to me before. This was one of those occasions . . . . Your consideration of the

> evidence should not be influenced in any way by any assumptions you may have made or any conclusions you may have drawn about delays.

There is no history of which we're aware of miscarriages of justice resulting because juries draw erroneous inferences from the replacement of a judge. See *United States v. Gayles*, 1 F.3d 735, 738 (8th Cir. 1993); *United States v. LaSorsa*, 480 F.2d 522, 531 (2d Cir. 1973).

The defendant complains that the new judge pressured the jury to complete its deliberations in a day and that with more time it might have acquitted him. There is no evidence to support that accusation of a judge noted for her patience. The first judge had assured the jury that the trial would not interfere with any of the jurors' vacation schedules. When trial resumed on August 2 the jury was down to 12 because one of the two alternates had been excused and the other had replaced a juror who had been excused. One of the remaining jurors had long-standing vacation plans for August 5, and the original judge had (with the government's consent) assured her when the government sought mandamus and the trial was adjourned that she would not need to show up on or after that date. When the trial resumed, another juror asked in open court what the jury should do in light of the possibility that the juror with vacation plans would leave before the trial ended. In response, and without objection by the defendant's lawyer, the judge said "we can't proceed" with fewer than 12 jurors. That was true (since the parties would not stipulate to a jury of 11, see Fed. R. Crim.

P. 23(b)(2)), though what was also true but she rightly did not say, because it would have sown confusion, is that while the *trial* could not continue without 12 jurors, if once the jury retired for its deliberations one of the jurors then decamped the judge could allow the remaining 11 to render a verdict even without the lawyers' consent. Fed. R. Crim. P. 23(b)(3).

August 4 turned out to be the last day of the trial. Closing arguments and the reading of the instructions to the jury took until the afternoon. The jury retired to consider its verdict at about 3:45 and returned 7 hours later with a verdict of guilty on eight counts and not guilty on the remaining six. The defendant argues that the jurors had rushed to complete their deliberations, knowing there would not be 12 jurors the next day. Given the strength of the government's case and the length of the jury's deliberations, and the fact that there was only one defendant and that the jury acquitted him on some counts, it is unlikely that even if they hadn't been expecting to lose the twelfth juror the next day, the jurors would have taken more time to deliberate than they did, though they might have broken at dinner time and resumed the following morning. The judge did not, as in the cases that the defendant cites to us, *United States v. Blitch*, 622 F.3d 658, 670 (7th Cir. 2010), and *United States v. Chaney*, 559 F.2d 1094, 1098 (7th Cir. 1977), set a deadline, either explicit or implicit, for the jury's deliberations. On the contrary, after instructing the jury, and only moments before the jury left the courtroom to deliberate, the judge told them: "I think I mentioned earlier that from this point on, the schedule is up to you. I realize [by] the way that the trial has been bumpy, and I will make every effort to accommodate your schedule from this moment on, whatever your decisions are. I appreciate your time. I think all of us do. You are excused to deliberate on your verdict." That was the opposite of pressuring the jury to complete its deliberations in a day. The jurors were unlikely to feel rushed when the judge had gone out of her way to tell them that she would make every effort to accommodate their schedules. Had the jurors been unable to agree on a verdict on August 4, the foreman would have told the judge that they couldn't reach a verdict and she would have either discharged them and declared a mistrial or allowed the 11 remaining jurors to return the next day and deliberate.

When the jury retired to deliberate, knowing that one juror would leave on vacation the next day and perhaps believing that 12 jurors had to be present to render a verdict, no juror asked the judge a question such as: "Does this mean we must render a verdict by the end of the day or can we just report our inability to reach a verdict?" Or: "What if we can't complete our deliberations by the end of the day?" Such questions would have flagged concerns that the judge would doubtless have addressed. No questions were asked. That suggests that the jurors were not concerned that the trial might end without a verdict unless they rushed their deliberations.

AFFIRMED.

COMMONWEALTH vs. TERRY L. PATTERSON.

445 Mass. 626

September 7, 2005 - December 27, 2005

Suffolk County

Present: MARSHAL., C.J., GREANEY, IRELAND, SPINA, COWIN, SOSMAN, & CORDY, JJ.

Evidence, Fingerprints, Scientific test, Expert opinion.

Discussion of the standard of review when considering a judge's decision regarding the admissibility of expert testimony based on scientific, technical, or other specialized knowledge. [639]

Statement of the analysis undertaken when considering the admissibility of expert testimony based on scientific, technical, or other specialized knowledge. [640-641]

This court concluded that a Superior Court judge hearing a criminal defendant's motion to suppress fingerprint evidence acted well within her discretion in ruling that latent fingerprint identification theory was generally accepted in the community of fingerprint examiners; that the analysis, comparison, evaluation, and verification (ACE-V) methodology used to compare a latent fingerprint impression to a fully inked fingerprint was generally accepted; and that the community of fingerprint examiners al-lowed enough room for and had enough debate to be considered a relevant technical community, and therefore, the judge had an adequate basis for concluding that ordinary single impression latent fingerprint identification is reliable [641-644]; however, this court concluded that the question of the reliability of ACE-V as applied to simultaneous impressions required separate analysis, and where the Commonwealth failed to carry its burden of proving general acceptance of the application of ACE-V to simultaneous impressions, where there was an absence of evidence of real testing of ACE-V as applied to simultaneous impressions, where the record contained no evidence of the promulgation of peer-reviewed standards relating to the application of ACE-V to simultaneous impressions, where there was no evidence pertaining to the error rate of ACE-V as applied to simultaneous impressions, and where there was a lack of accepted explicit universal standards controlling the application of ACE-V to simultaneous impressions, this court remanded the case for further proceedings [644-654].

INDICTMENTS found and returned in the Superior Court Department on October 27, 1993.

After review by this court, 432 Mass. 767 (2000), a pretrial motion to suppress evidence was heard by Margaret R. Hinkle, J., and a question of law was reported by her to the Appeals Court.

Page 627

The Supreme Judicial Court granted an application for direct appellate review.

John H. Cunha, Jr. (Helen Holcomb & Charles Allan Hope with him) for the defendant.

Donna Jalbert Patalano, Assistant District Attorney, for the Commonwealth.

The following submitted briefs for amici curiae:

Robert C. Cosgrove, Assistant District Attorney, for District Attorney for the Berkshire District & others.

David M. Siegel, Stanley Z. Fisher, & Daniel Givelber for

New England Innocence Project & others.

Lisa J. Steele for National Association of Criminal Defense Lawyers & others.

LaDonna J. Hatton & Christopher Pohl, Special Assistant Attorneys General, for Secretary of Public Safety.

CORDY, J. In 1995, Terry L. Patterson was convicted of the murder of a Boston police detective. [Note 1] His conviction was based in large part on the expert testimony of a member of the Boston police latent fingerprint section, who used the most common method of latent fingerprint identification, ACE-V, [Note 2] to determine that four latent impressions found on the victim's vehicle were left by Patterson. While no single latent impression, on its own, could reliably be matched to its allegedly corresponding finger, the fingerprint examiner based his testimony on the cumulative similarities observed between the impressions and their corresponding fingers. The examiner opined that the four impressions could be analyzed collectively because he believed them to be simultaneous impressions, that is, impressions of multiple fingers made by the same hand at the same time.

After this court set aside Patterson's convictions on a ground not relevant to this appeal, see Commonwealth v. Patterson, 432 Mass. 767 , 768 (2000), Patterson moved to exclude all fingerprint evidence from his retrial because, in his view, the Commonwealth's latent fingerprint identification evidence was

Page 628

unreliable and thus inadmissible under Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993) (Daubert), and Commonwealth v. Lanigan, 419 Mass. 15 (1994) (Lanigan). After conducting an evidentiary hearing, a Superior Court judge denied Patterson's motion and reported the issue to the Appeals Court. We granted Patterson's application for direct appellate review to determine whether the judge abused her discretion in finding that the Commonwealth had established the reliability of its latent fingerprint identification evidence.

Consistent with the decisions of other courts that have considered the issue since Daubert, we conclude that the underlying theory and process of latent fingerprint identification, and the ACE-V method in particular, are sufficiently reliable to admit expert opinion testimony regarding the matching of a latent impression with a full fingerprint. In this case, however, the Commonwealth needed to establish more than the general reliability of latent fingerprint identification. It needed to establish that the theory, process, and method of latent fingerprint identification could be applied reliably to simultaneous impressions not capable of being individually matched to any of the fingers that supposedly made them. On the record before the judge below, the Commonwealth failed to meet its burden. [Note 3]

1. Background. Before addressing the legal claims, we will briefly lay out the theory behind and modern application of latent fingerprint identification as well as the factual history of this case. We rely principally on the findings of fact made by the motion judge in connection with Patterson's motion to exclude the fingerprint evidence and on the transcript of the previous trial.

a. Latent fingerprint identification theory. Fingerprint

Page 629

evidence has been used extensively in criminal investigations and trials for more than one hundred years. Fingerprints are left by the deposit of oil on contact between a surface and the friction ridges of a finger. Latent fingerprints are fingerprint impressions that are not visible to the naked eye without chemical enhancement. These latent print impressions are almost always partial and may be distorted due to less than full, static contact with the object and to debris covering or altering the latent impression.

The theory behind latent fingerprint identification, called "individualization," is that a positive identification can result from the comparison of two fingerprints containing sufficient quality and quantity of detail. The underlying premise of this theory is the uniqueness and permanence of human friction ridge arrangements - that no two fingers, even on the same hand of the same person, contain the same ridge pattern. This uniqueness begins during prenatal development, when a template of the ridge patterns appears on the skin, and absent damage to the template, remains in the same exact form throughout one's life. A fingerprint should accordingly only match one finger of one person in the world.

b. The process of identification (ACE-V). The uniqueness of two full fingerprints does not, in and of itself, prove that one small portion of a fingerprint cannot mirror one portion of another fingerprint. And because latent print impressions left at crime scenes are often partial impressions of a full fingerprint, subject to significant distortions, it is a question of significant dispute as to how much detail in the latent print must be demonstrable to assert reliably its identity with a known fingerprint. Consequently, law enforcement and forensic scientists have endeavored to create and refine the method by which they identify the true "owner" of latent print impressions. A latent fingerprint impression lifted from a crime scene is compared to a full exemplar print taken from the suspect under controlled circumstances by dipping his finger in ink and slowly impressing his entire finger on a card in order to ensure full transcription of the finger. Matches of a latent print to a full print are made in several ways. A latent print can be processed through a computerized system that compares it to a rather

Page 630

large database of known full prints. Alternatively, a set of a suspect's fully inked fingerprints can be given to an examiner for comparison purposes. Either way, the fingerprint examiner ultimately compares the latent print to its potentially matching full print using a method known as ACE-V (analysis, comparison, evaluation, and verification). [Note 4]

In the analysis stage of ACE-V, the examiner looks at the first of three levels of detail ("level one") on the latent print. Level one detail involves the general ridge flow of a fingerprint, that is, the pattern of loops, arches, and whorls visible to the naked eye. The examiner compares this information to the exemplar print in an attempt to exclude a print that has very clear dissimilarities. At this stage, the examiner also looks for focal points - or points of interest - on the latent print that could help prove or disprove a match. Such focal points are often at the boundaries between different ridges in the print. The examiner will then collect level two and level three detail information about the focal points he has observed. Level two details include ridge characteristics (or Galton Points) like islands, dots, and forks, formed as the ridges begin, end, join or bifurcate. Level three details involve microscopic ridge at-tributes such as the width of a ridge, the shape of its edge, or the presence of a sweat pore near a particular ridge.

In the comparison stage, the examiner compares the level one, two, and three details of the focal points found on the latent print with the full print, paying attention to each characteristic's location, type, direction, and relationship to one another. The comparison step is a somewhat objective process, as the examiner simply adds up and records the quantity and quality of similarities he sees between the prints. In the evaluation stage, by contrast, the examiner relies on his subjective judgment to determine whether the quality and quantity of those similarities are sufficient to make an identification, an exclusion, or neither.

Page 631

While some jurisdictions require (or used to require) a minimum number of Galton point similarities to declare an individual match between a latent and full print, most agencies in the United States no longer mandate any specific number. [Note 5] Rather, the examiner uses his expertise, experience, and training to make a final determination. There is a rule of examination, the "one-discrepancy" rule, that provides that a non identification finding should be made if a single discrepancy exists. However, the examiner has the discretion to ignore a possible discrepancy if he concludes, based on his experience and the application of various factors, that the discrepancy might have been caused by distortions of the fingerprint at the time it was made or at the time it was collected.

Assuming a positive identification is made by the first examiner, the verification step of the process involves a second

Page 632

examiner, who knows that a preliminary match has been made and who knows the identity of the suspect, repeating the first three steps of the process.

c. Simultaneous impressions. The ACE-V method is usually employed to match one latent fingerprint impression to one fully inked fingerprint. Often, however, a person leaves latent impressions of multiple fingers on objects that he touches. Such fingerprint impressions left by the same person at the same time are referred to as simultaneous impressions. A difficulty arises when no single latent impression in the cluster of simultaneous impressions has a sufficient quantity or quality of similar detail to be matched reliably to a single fully inked fingerprint using the ACE-V approach. In such cases, some fingerprint examiners have applied the ACE-V method to identify suspects based on the aggregate number of similarities between latent and full impressions of multiple fingers.

For example, assume five latent fingerprint impressions are found on a table in a manner that suggests they were left by a person placing his full hand down on that table. If each of those prints had only three points of similarity of moderate quality relative to a corresponding fully inked fingerprint, an examiner who requires eight similarity points of moderate detail to make an identification would not be able to match any individual impression to any individual fully inked fingerprint. If the examiner applied ACE-V collectively to the simultaneous impressions, however, he might conclude that fifteen points of similarity (five fingers with three similarity points per finger) between the impressions left on the table and a suspect's hand signifies a definite match. A fingerprint examiner would first have to use his expertise, experience, and training to determine whether the several latent impressions were in fact created simultaneously. In doing so, the examiner apparently may take into account the distance separating the latent impressions, the orientation of the impressions, the pressure used to make the impression, and any other facts the examiner deems relevant. The record does not, however, indicate that there is any approved standardized method for making the determination that two or more print impressions have been made simultaneously.

d. Factual history. On September 26, 1993, the body of

Page 633

Detective John Mulligan of the Boston police department was found in his truck outside a Walgreens store in the Roslindale section of Boston. Detective Mulligan, who performed paid security work for the store, had been shot five times at extremely close range. Detective Mulligan's department-issued sidearm was missing from his holster. A store employee saw Detective Mulligan asleep in his truck at 3:30 A.M. and found him dead with a bloodied face fifteen minutes later. Several witnesses recalled seeing two black men near the Walgreens in the early morning of September 26, but none could offer more than a very vague and general description of those men. On October 27, 1993, Patterson was indicted for armed robbery, two counts of possession of a dangerous weapon, and the murder of Detective Mulligan. At trial, the Commonwealth argued that two men, one of whom was Patterson, happened on the sleeping detective and seized the opportunity to rob him of his firearm.

The Commonwealth introduced evidence, through the testimony of Robert Foilb of the Boston police latent fingerprint section, that four latent fingerprint impressions recovered from the window of the driver's door of Detective Mulligan's truck were made simultaneously by four fingers on Patterson's left hand as he closed the driver's door. Foilb explained that the location of the print impressions in relation to each other and the direction and manner in which they each streaked on the glass reinforced his belief that they were left by multiple fingers of the same hand at the same time. Comparing these prints to inked fingerprints of all ten of Patterson's fingers, Foilb concluded that the four prints corresponded to the little finger, ring finger, middle finger, and index finger of Patterson's left hand.

On cross-examination, Foilb testified that the locally accepted norm for successfully matching a latent print to a full fingerprint was eight points of similarity. He conceded that none of the four latent fingerprints contained enough similarity with the fully inked print of Patterson's corresponding finger to satisfy this generally accepted minimum norm and be individually matched. Notwithstanding this concession, Foilb opined that he could conclusively determine that the four simultaneous print impressions were those of Patterson because the sum of the

Page 634

points of similarity on the four fingers, which he determined respectively to be six, five, two, and zero, totaled thirteen - thus exceeding the eight point similarity standard. He testified that the similarities found in simultaneous impressions "can be counted as a total number because there is no other way to have those fingerprints put on an object."

On February 1, 1995, a jury convicted Patterson on all charges. On December 6, 2000, his convictions were reversed because of a conflict of interest that deprived Patterson of the effective assistance of counsel at trial. See Commonwealth v. Patterson, 432 Mass. 767 , 781 (2000). Noting that "the evidence was sufficient to support the convictions," id. at 768, this court remanded the case to Superior Court for a new trial. See id. at 781.

On October 11, 2002, Patterson filed a motion in limine to exclude all fingerprint evidence from his retrial. The Commonwealth sought to offer latent fingerprint identification evidence similar to the evidence it presented at Patterson's first trial for the purpose of placing him at the scene of the crime, including testimony regarding the four supposedly simultaneous impressions. [Note 6] Patterson argued that the Commonwealth's latent fingerprint identification failed to meet the Daubert reliability standard governing the admissibility of expert opinion testimony that this court adopted in Lanigan, supra. He also contended that the

ACE-V methodology was an unreliable application of the theory of latent fingerprint identification. Finally, he argued that even if the judge found the ACE-V process to be reliable in matching one particular latent print of a finger to a fully inked print of the same finger, this method was still unreliable when applied to cases of simultaneous impressions in which none of the individual prints could be separately matched. In May, 2004, a judge in the Superior Court held a Daubert-Lanigan hearing over five days, which included live testimony from two witnesses, affidavits, previously recorded expert testimony from

Page 635

other cases, and a large number of exhibits. [Note 7] The Commonwealth offered live testimony of Supervisory Fingerprint Specialist Stephen Meagher of the Federal Bureau of Investigation (FBI) and transcripts of testimony given at a similar hearing, see United States v. Mitchell, 365 F.3d 215 (3d Cir. 2004), by Royal Canadian Mounted Police Fingerprint Examiner David R. Ashbaugh and noted biological anthropologist Dr. William J. Babler. Patterson offered the live testimony of George Washington University professor of forensic sciences James E. Starrs and the transcripts of testimony of David Stoney, a doctor of forensic sciences from the University of California at Berkeley, and Simon A. Cole, who has a doctorate in science and technology from Cornell University, also given in the Mitchell case.

e. The judge's original order. On October 12, 2004, the motion judge issued a detailed order (original order) denying Patter-son's motion. In her order, the judge explained that the admissibility of expert testimony depended on the reliability of the theory and methodology that the expert used to reach an opinion. The judge relied on the test established in Daubert to determine the reliability of latent fingerprint identification theory and the ACE-V methodology. While recognizing that the test for reliability was flexible and did not necessarily require an examination of all or even most of the five factors that Daubert recognized as potentially relevant to such an inquiry, she carefully and thoroughly applied each factor to latent fingerprint identification theory and to the ACE-V methodology. Those factors are: (1) whether the testimony's underlying theory and application is generally accepted in the relevant scientific-technical community, (2) whether the theory and application have been or can be subjected to testing, (3) whether they have

Page 636

been subjected to peer review and publication, (4) whether the application has an unacceptably high known or potential error rate, and (5) whether the application is governed by recognized standards.

The judge first concluded that both latent fingerprint identification theory in general and the ACE-V methodology in particular are generally accepted in the fingerprint examiner community. [Note 8] For support, she pointed to an FBI survey of fifty-three domestic and foreign jurisdictions that confirmed the unanimous and long-standing acceptance of latent fingerprint identification theory. Similarly, she found that there existed "overwhelming support for ACE-V in the forensic identification community." She rejected Patterson's contention that the community was not broad enough to count as a relevant community for Daubert purposes. Finally, the judge noted the long and virtually universal history of court acceptance of fingerprint identification evidence.

Despite recognizing that Lanigan posited that "general acceptance" would often be the only factor necessary to the inquiry, the judge proceeded to address the other Daubert factors, beginning with the testability of latent fingerprint identification theory and the ACE-V methodology. She concluded that, although the theory underlying latent fingerprint identification - uniqueness and permanence of fingerprints - is testable and has been successfully tested, the notion that a person can be positively identified from an individual latent fingerprint impression that contains sufficient quantity and quality of ridge detail is somewhat less testable. The judge noted that ACE-V defies easy testing because it does not require a minimum number of similarities, but rather operates on a subjec-

Page 637

tive sliding scale. The judge also explained that an FBI study that matched 50,000 simulated latent prints against 50,000 full prints with no false positives was not particularly helpful. She considered it to be a flawed test because the simulated latent prints were not subject to real world distortions. The judge ultimately discounted the testability problem, however, because the print matches used in court are usually accompanied by expert testimony that establishes the number of similarities observed between the latent and full print and can form the basis of testing by the opposition's independent examiner.

The judge next concluded that the ACE-V methodology had been subjected to limited peer review in forensic publications and during the process of formalizing the ACE-V guidelines by the Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST). [Note 9] However, the judge disagreed with the Commonwealth that the verification procedure, which she described as illusory, constituted peer review, and thus found that this factor only slightly favored admission of the evidence.

Turning to error rate, the judge found the ACE-V error rate for false positive identifications to be very low. Relying largely on an extensive FBI survey in which no State agency returned a false positive when attempting to match a set of latent prints against seventy million full ten-prints records, the judge noted that recent high profile cases of misidentification do not alter the over-all low error rate. She similarly discounted evidence of poor scores that examiners sometimes receive on routine proficiency tests administered by their respective agencies.

Finally, the judge concluded that ACE-V is controlled by ap-

Page 638

propriate standards despite the lack of a uniform, minimum number of similarities requisite to declaring a match. The judge determined that the rigorous qualification and training standards normally required for FBI fingerprint examiners help to control the operation of ACE-V, which itself is a relatively uniform procedure. With the Daubert factors generally favoring admission, the judge denied Patterson's motion to exclude the Commonwealth's fingerprint evidence identifying Patterson as the person who left four simultaneous impressions on the door of the victim's truck.

f. The judge's supplemental order. On November 29, 2004, the judge issued a four-page supplemental order, in which she acknowledged that her previous decision "did not explicitly ad-dress the reliability of the process of making an identification based on `simultaneous' impressions." She concluded that the application of ACE-V to simultaneous impressions was sufficiently reliable to be admitted. This conclusion was based primarily on Agent Meagher's testimony that the use of simultaneous impressions positively to identify a person "involves the exact process involved in individualizing a single latent print, simply applying ACE-V to the composite of level one, two, and three detail of multiple prints from the same hand." The judge added that the use of a fingerprint examiner's judgment to determine whether multiple impressions were deposited simultaneously does not make the process unreliable. The judge also noted that Agent Meagher testified that the use of ACE-V in situations of simultaneous impressions was generally accepted in the community of fingerprint examiners, and found that fingerprint examiners in Great Britain sometimes use a similar approach. The judge acknowledged, however, that one of the Commonwealth's witnesses, Royal Canadian Mounted Police Fingerprint Examiner David Ashbaugh, was of the view that the application of ACE-V to simultaneous impressions is a "hodgepodge" approach not based on any science. Finally, the judge found that the absence of a specific peer-reviewed study or published article validating the use of ACE-V in situations of simultaneous impressions is not fatal to its admissibility, as reliability can be shown through other Daubert factors.

To the extent that the judge performed a separate factor-by-

Page 639

factor Daubert analysis of the application of ACE-V to simultaneous impressions, her analysis was implied and brief. This is apparent when her supplemental order is compared to her painstaking review of the record in her original order. The most plausible reading of the judge's supplemental order is that she found the application of ACE-V to simultaneous impressions sufficiently reliable because she concluded that it did not differ in any significant way from the use of ACE-V to match a single latent fingerprint impression.

g. Reservation and report. On January 14, 2005, the judge reserved and reported her orders to the Appeals Court because, in her view, the question of admissibility of fingerprint evidence in this case is both important and doubtful. See Mass. R. Crim. P. 34, 378 Mass. 905 (1979).

2. Discussion. Trial judges serve a gatekeeper function with respect to expert opinion testimony based on specialized knowledge. See Lanigan, supra at 26. "If the process or theory underlying [an] . . . expert's opinion lacks reliability, that opinion should not reach the trier of fact." [Note 10] Id.

a. Standard of review. We review a judge's Lanigan decision for abuse of discretion. See General Elec. Co. v. Joiner, 522 U.S. 136, 141-143 (1997); Canavan's Case, 432 Mass. 304 , 311-312 (2000). While our review under this standard is deferential and limited, it is not perfunctory. A judge's findings must apply the correct legal standard to the facts of the case and must be supported by an examination of the record. See id. at 312 ("applying an abuse of discretion standard on appellate review will allow trial judges the needed discretion to conduct the inherently fact-intensive and flexible Lanigan analysis, while preserving a sufficient degree of appellate review to as-sure that Lanigan determinations are consistent with the law and supported by a sufficient factual basis in the particular case").

Page 640

b. The Lanigan analysis. In Daubert, the United States Supreme Court announced a new test to govern the admissibility in Federal courts of expert testimony based on scientific, technical, and other specialized knowledge. [Note 11] While the most common pre-Daubert test, set forth in Frye v. United States, 293 F. 1013 (D.C. Cir. 1923) (Frye), required the theory and methodology in question to be generally accepted by a relevant scientific community, the new, more flexible standard laid out five factors that a court might consider in a determination of reliability in the totality of the circumstances. Daubert, supra at 593-594. The Frye general acceptance test became simply one of the five factors in the Daubert test. See Lanigan, supra at 25 ("In its Daubert opinion, the Court recognized that general acceptance . . . was a relevant factor in determining . . . admissibility . . . . But such acceptance, the essential ingredient of the Frye principle, is not the sole test" [citation omitted]).

Massachusetts historically hewed to the Frye general acceptance test. See Lanigan, supra at 24, quoting Commonwealth v. Cumin, 409 Mass. 218 , 222 (1991) ("Our test . . . has usually been . . . `whether the community of scientists involved generally accepts the theory or process' "). In Lanigan, supra at 25-26, however, we adopted, in part, the new Daubert standard. In so doing, we cautioned that "general acceptance in the relevant . . . community will continue to be the significant, and often the only, issue." Id. at 26.

Lanigan's progeny make clear that general acceptance in the relevant community of the theory and process on which an expert's testimony is based, on its own, continues to be sufficient to establish the requisite reliability for admission in Massachusetts courts regardless of other Daubert factors. See Commonwealth v. Sands, 424 Mass. 184 , 185-186 (1997) ("party seeking to introduce scientific evidence may lay a foundation either by showing that the underlying scientific theory is generally accepted within the relevant scientific community, or by

Page 641

showing that the theory is reliable or valid through other means" [emphasis added]); Canavan's Case, supra at 310 (Lanigan's partial adoption of Daubert was merely "to account for [the] circumstance" where "strict adherence to the Frye test" caused otherwise reliable evidence to be excluded because it had not yet become generally accepted). See also Commonwealth v. Senior, 433 Mass. 453 , 458-459 (2001) (same). Where general acceptance is not established by the party offering the expert testimony, a full Daubert analysis provides an alternate method of establishing reliability. See Lanigan, supra at 26 ("proponent of scientific opinion evidence may demonstrate the reliability or validity of the underlying scientific theory or process by some other means, that is, without establishing general acceptance"). See also Commonwealth v. Sands, supra at 186 n.1 (the absence of general acceptance is simply "a factor for the court to consider" in its subsequent Daubert analysis).

c. Reliability of latent fingerprint identification and ACE-V in general. The judge acted well within her discretion in concluding that latent fingerprint identification theory is generally accepted in the community of fingerprint examiners. [Note 12] At the hearing, the Commonwealth presented a 1999 survey, conducted by Agent Meagher, confirming that the top law enforcement agencies in all fifty States, the District of Columbia, Canada, and England accept the theory of latent fingerprint identification. Each jurisdiction reported that it accepted the use of both fully

Page 642

recorded and latent fingerprints as a positive means of identification. This survey is a sufficient basis on which the judge could have concluded there to be general acceptance of the theory in the fingerprint examiner community. See United States v. Mitchell, 365 F.3d 215, 241 (3d Cir. 2004) ("The answer" to whether fingerprint identification is generally accepted in forensic community "is yes, as demonstrated by the results of the FBI's survey of state agencies"). Other evidence before the judge, and findings reported in other cases, reveal that fingerprint experts from countries around the world also accept and apply the theory of latent fingerprint identification. See id. at 222; United States v. Llera Plaza, 188 F. Supp. 2d. 549, 555, 566-567 (E.D. Pa. 2002).

The judge's additional conclusion that the ACE-V methodology used to compare a latent fingerprint impression to a fully inked fingerprint is generally accepted is also adequately supported by the record. Agent Meagher represented, and it was not disputed, that ACE-V is the standard methodology used throughout the United States and other parts of the world. Supporting Agent Meagher's testimony, the record establishes that SWGFAST has, after multiple levels of debate and peer review by its own members and the International Association for Identification (IAI), adopted and published fingerprint identification standards setting forth the ACE-V methodology.

Notably, Patterson does not dispute the assertion that the fingerprint examiner community generally accepts either latent fingerprint identification theory or the ACE-V methodology. Rather, he argues that this community is not sufficiently broad to constitute "a relevant scientific community" for purposes of gauging general acceptance, and that the problem is acute because the fingerprint examiner community lacks financially disinterested academics and is prone to stifling dissent.

Given that Lanigan applies to technical evidence as well as scientific evidence, Canavan's Case, 432 Mass. 304 , 313 (2000), we do not concern ourselves with whether fingerprint examiners are scientists or technicians. See United States v. Mitchell, supra at 241 (rejecting argument that "there is no scientific community that generally accepts fingerprint identification" because "the scientific/nonscientific distinction

Page 643

is irrelevant after [Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999)]"); United States v. Llera Plaza, supra at 563, quoting Fed. R. Evid. 702 ("the fingerprint community's `general acceptance' of ACE-V should not be discounted because finger-print specialists . . . have `technical . . . knowledge' . . . rather than `scientific . . . knowledge' . . . and hence are not members of what Daubert termed a scientific community"). A technical community, or a community of experts who have some other specialized knowledge, can qualify as a relevant Daubert community in the same way a scientific community can.

We are, however, cognizant of the need to define the relevant community. In Canavan's Case, this court explained that "[a] relevant scientific community must be defined broadly enough to include a sufficiently broad sample of scientists so that the possibility of disagreement exists," and we cautioned trial judges not to "define the `relevant scientific community' so narrowly that the expert's opinion will inevitably be considered generally accepted." Id. at 314 n.6. In the context of technical forensic evidence, the community must be sufficiently broad to permit the potential for dissent.

The judge properly ensured that the technical community in which latent fingerprint identification and ACE-V is generally accepted is broad enough to include "some practitioners who acknowledge flaws in the methodology" and tolerant enough to allow "some, albeit, limited room for dissent." For example, the guidelines and standards developed by SWGFAST commit-tees are subject to repeated discussion, critique, and debate by the entire SWGFAST community and by members of the IAI. Additional room for disagreement lies in the ongoing debate over how many points of similarity, if any, are needed to conclusively make a match. See United States v. Llera Plaza, supra at 567. Further, as we will discuss below, even one of the fingerprint examiners whose testimony in the Mitchell case was offered by the Commonwealth - David Ashbaugh - registered objection to the particular application of ACE-V that the Commonwealth is seeking to use in this case.

We cannot conclude that the judge abused her discretion in finding that the community allowed enough room for and had

Page 644

enough debate to be considered a relevant technical community. This finding has support in the unanimity with which modern courts have concluded that latent fingerprint identification theory and ACE-V are generally accepted by the fingerprint examiner community and that that community qualifies as a relevant community for Daubert purposes. Cf. Commonwealth v. Senior, 433 Mass. 453 , 461-462 (2001) ("Given the evidence of reliability presented by the Commonwealth, and the admissibility of similar evidence in other jurisdictions, the judge did not abuse his discretion in finding that the evidence . . . was sufficiently reliable"). See, e.g., United States v. Mitchell, supra at 241 (concluding fingerprint identification is generally accepted in forensic community); United States v. Crisp, 324 F.3d 261, 268-269 (4th Cir. 2002), cert. denied, 540 U.S. 888 (2003), quoting United States v. Starzecpyzel, 880 F. Supp. 1027, 1038 (S.D.N.Y. 1995) (rejecting claim that general acceptance of fingerprint identification in expert community should be discounted because community "is devoid of financially disinterested parties such as academics"); United States v. Sullivan, 246 F. Supp. 2d 700, 702-703 (E.D. Ky. 2003) ("ACE-V methodology easily satisfies the general acceptance factor of Daubert. . . . Daubert requires the court to consider whether ACE-V has been accepted by a substantial portion of the pertinent scientific or technical community"); United States v. Llera Plaza, supra at 563-564 (accepting as relevant community fingerprint specialists with technical or specialized knowledge). The defendant has given us no reason to reject the common judicial wisdom that considers the fingerprint examiner community a relevant community for Daubert purposes.

Because both latent fingerprint identification theory and the use of ACE-V to match a latent impression to a fully inked fingerprint are generally accepted by a sufficiently broad community of technical experts, the judge had an adequate basis for concluding that ordinary single impression latent fingerprint identification is reliable, and did not need to examine the other Daubert factors. Her original order is affirmed.

d. Reliability of latent fingerprint impression (and ACE-V) applied to simultaneous impressions. While establishing that the reliability of latent fingerprint identification and ACE-V is

Page 645

necessary to admit the Commonwealth's fingerprint evidence, it is not sufficient. Rather, the evidence can only be admitted if, in addition to the reliability of the theory and process in general, the process is reliable when applied to the specific issue about which the expert is proposing to testify. See Kumho Tire Co. v. Carmichael, supra at 153-154 ("specific issue before the court was not the reasonableness *in general* of [the expert's method]. Rather, it was the reasonableness of using such an approach . . . to draw a conclusion regarding *the particular matter to which the expert testimony was directly relevant*"); Canavan's Case, supra at 311-312, quoting State v. Alberico, 116 N.M. 156, 169 (1993) (rejecting de novo standard of review in Lanigan cases because that standard incorrectly assumes that "application of a particular scientific method would not vary from case to case and thus would be worthy of a judicial stamp of approval or rejection as a matter of law").

In this case, the Commonwealth proposes to call a State trooper, trained in fingerprint examination, to testify that he used the ACE-V methodology positively to identify Patterson as the person who left four latent simultaneous impressions on the victim's truck, despite the fact that the application of ACE-V to any of the individual latent impressions would not have led to a match. Such testimony is based on the theory that once a group of latent impressions are identified as simultaneous impressions, an otherwise unacceptably small number of similarities between each of the impressions and its allegedly corresponding fully inked fingerprint can form the basis for a collective determination as to whether the entire group of latent impressions matches a corresponding group of full fingerprints. To gain admission of the trooper's testimony, then, the Commonwealth must establish that adding up similarity points of simultaneous impressions is a reliable way to use ACE-V to effectuate latent fingerprint identification.

Instead of engaging in the deliberate factor-by-factor analysis that she undertook with respect to the more general theory and application of latent fingerprint identification, however, the judge assumed the reliability of the application of ACE-V to simultaneous impressions because it "involves the exact process" simply applied to a "composite" record of the detail

Page 646

of "multiple prints from the same hand." In doing so, the judge appeared to endorse the Commonwealth's position that any added potential for error in the identification process caused by applying ACE-V to simultaneous impressions is relevant only to the weight of evidence and not its reliability and, therefore, is not relevant to its admissibility.

In support of the judge's conclusion, the Commonwealth points to our recent decision in Commonwealth v. Gaynor, 443 Mass. 245 (2005), upholding the admission of deoxyribonucleic acid (DNA) evidence. [Note 13] In that case, we suggested that the defendant's arguments attacking the reliability of particular DNA tests (regarding "mixed sample" testing) failed in part because "the issues raised by the defendant went to the weight of the evidence, not its admissibility." Id. at 267. The Commonwealth seizes on these statements to argue that once ACE-V is found to be generally reliable, any application of it also must be sufficiently reliable to be admitted in evidence.

The Commonwealth has misread Commonwealth v. Gaynor, supra. Our comments presupposed a finding that the particular application of the DNA test used was reliable. See id. at 265 ("The judge's findings that [Cellmark Diagnostic Laboratories'] methodology in reporting tests of a mixed sample with an identifiable primary contributor in the same way it reports tests of a single source sample conforms to the recommendations of the [National Resource Council], and that Cellmark's methodology in dealing with the presence of mixtures or technical artifacts is generally accepted within the scientific community, were made with record support and well within his discretion" [emphasis added]). The opinion noted with emphasis that the reliability of mixed sample testing was considered reliable because it was generally accepted in its own right, not by mere virtue of the reliability of single sample testing. [Note 14]

Consistent with our opinion in Commonwealth v. Gaynor, su-

Page 647

pra, we recognize that applying ACE-V to simultaneous impressions sufficiently alters the process to require its own reliability inquiry. An examiner must first determine whether the impressions were simultaneously left at the scene, and then apply ACE-V not to a single finger but to multiple separate sections of a whole hand. Notwithstanding contentions that the enlargement of the zone of comparison does not change the process, the process is fundamentally altered when an examiner is asked to make a determination, as he was in this case, that a particular latent impression matches a particular full fingerprint despite the absence of enough similarity to determine a match solely by comparing those two prints.

Arguments that the application of ACE-V to simultaneous impressions is simply an extension of ACE-V in general prove too much. Under this theory, any impressions - whether simultaneous or not - that an examiner believed to be left by the same person could be subject to ACE-V testing free from a Daubert inquiry. Likewise, because each full hand print is apparently unique, it would follow that, under this theory, an examiner could subject impressions from two different hands (that did not contain enough similarities on their own to declare a match) to this cumulative analysis without requiring a separate Daubert inquiry. [Note 15]

In Commonwealth v. Gaynor, supra, we did not hold that judges should abdicate their role in reviewing various applications of generally accepted methodologies on which experts base their opinions. To the contrary, we noted that the "determination of the reliability of the testing process entails a fact-based inquiry, including questions of credibility." Id. at 264. While questions of credibility are traditionally left for the

Page 648

jury, we explained that, in this context, this inquiry was the responsibility of the judge. See id., citing Commonwealth v. McNickles, 434 Mass. 839 , 850 (2001) ("analysis calls on a judge to determine whether testing was properly performed").

It is beyond doubt that Daubert and Lanigan envision that the jury will decide the ultimate question of the conclusiveness of the results of a reliable application of a methodology, see Daubert, supra at 594-595 ("The focus [of the Daubert inquiry], of course, must be solely on principles and methodology, not on the conclusions that they generate"), because even reliable procedures can lead to incorrect results. A judge's evidentiary determination that a particular application is reliable simply al-lows the jury to determine whether such an application led to a reliable result, taking into account all of the facts at hand. Judges, however, need not admit (and juries need not wrestle with) every application of a testing method - no matter how dubious - merely because another application of the method has been deemed reliable. See Commonwealth v. McNickles, supra (accepting distinction between reliability of general method of testing DNA and reliability of particular application of that test). See also Commonwealth v. Cumin, 409 Mass. 218 , 222 n.7 (1991) ("Future challenges should focus on the soundness . . . of the particular testing process . . . and, if raised, on the proper implementation of that process in the given case"). Otherwise, the traditional role of judges as gatekeepers - protecting juries from evidence that had little chance of being reliable - would be significantly and needlessly diminished. See Commonwealth v. McNickles, supra at 850 ("judge's gate-keeper role under Commonwealth v. Lanigan, supra, includes the obligation to determine whether the testing at issue was conducted properly [and not just whether the testing method is theoretically reliable]").

In sum, the procedure that we adopted in Lanigan includes ensuring not only the reliability of the abstract theory and process underlying an expert's opinion, but the particular application of that process. The question of the reliability of ACE-V as applied to single latent impressions is distinct from the question of the reliability of ACE-V as applied to simultaneous impressions. The application of ACE-V to simultaneous

Page 649

impressions must therefore be subjected to its own Daubert analysis. We now proceed with that analysis.

i. General acceptance of applying ACE-V to simultaneous impressions. As we have explained, if the Commonwealth establishes that the application of ACE-V to simultaneous impressions is generally accepted in the fingerprint examiner community, the evidence is properly admitted. The judge found that "according to [Agent] Meagher, the use of ACE-V . . . to make individualization determinations from simultaneous impressions is generally accepted in the community of qualified fingerprint examiners." Unlike his testimony in the single impression context, however, Agent Meagher's testimony is conclusory and unsupported by any evidence, let alone an extensive multi-jurisdictional survey. The Commonwealth did not present evidence that any domestic agency or jurisdiction - save for the now disbanded Boston police fingerprint unit and the State police - relies on simultaneous impressions for identification purposes. Likewise, with the exception of Great Britain, there is no evidence in the record that any foreign jurisdiction applies ACE-V to simultaneous impressions. With regard to Great Britain, one of the Commonwealth's own experts, David Ashbaugh, a noted fingerprint examiner, described its use as a "weird doctrine." [Note 16] He further explained that the application of ACE-V to simultaneous impressions was one of several uses of ACE-V in England that "has resulted in a hodgepodge of doctrine that is far removed from the truth." Ashbaugh suggests that this application of ACE-V "require[s] a certain leap of faith" and has "no supporting rationale."

Moreover, the Commonwealth did not present evidence that SWGFAST, IAI, or any other fingerprint examination society accepts or recommends the application of ACE-V to simultaneous impressions. At best, the record lacks evidence of widespread acceptance of the application of ACE-V to simultaneous impressions by the fingerprint examiner com-

Page 650

munity. [Note 17] Because the Commonwealth has failed to carry its burden of proving general acceptance, we turn to the four remaining Daubert factors.

ii. Testing. We judge this factor by inquiring whether this application of ACE-V can be or has been tested. The judge's supplemental order noted that no specific study or scientific article has validated the application of ACE-V to simultaneous impressions. Agent Meagher testified that he was "not aware of any studies that have been performed to validate the application of ACE-V to simultaneous impressions to make an identification." Neither is this court.

In her original order, the judge explained her concern that the subjectivity involved in the ACE-V process means the process itself defies easy testing. Such concerns are exacerbated when simultaneous impressions are involved. There are presently no formalized standards governing an examiner's determination that impressions have been simultaneously made, leaving that determination largely to the judgment of the examiner. [Note 18]

The judge explained, however, that most important to her determination of the potential testability of latent fingerprint

Page 651

identification was that the one-discrepancy rule makes even subjective evaluations testable because any single discrepancy is enough to disprove a match. The one-discrepancy rule is, unfortunately, less than it appears. Fingerprint examiners can and often do ignore one or more discrepancies in a match or in simultaneous impression matches. They do so by reasoning that the discrepancy was created by a distortion or unnatural alteration of the impression.

The judge also noted that any particular result of the ACE-V process is testable by virtue of the in-court adversary process. That is, an independent examiner can challenge the conclusion of the Commonwealth's expert based on the specific criteria articulated by that expert used to declare a match. See United States v. Harvard, 260 F.3d 597, 601 (7th Cir. 2001). However, adversary testing is not what the Supreme Court meant when it discussed testing as an admissibility factor. See Commonwealth v. Cumin, 409 Mass. 218 , 222 n.7 (1991), citing United States v. Two Bulls, 918 F.2d 56, 61 (8th Cir. 1990) (until questions of reliability are determined by judge, "jury should not be given the evidence and allowed to determine the validity and soundness of the process because evidence of this character has too great a potential for affecting a jury's judgment"). Concluding that a test is reliable merely because testimony based on its results can be cross-examined in front of a jury puts the cart before the horse. In the absence of any real testing of ACE-V, at least as applied to simultaneous impressions, we conclude that this factor favors exclusion.

iii. Peer review and publication. In her original order regarding latent fingerprint identification, the judge correctly concluded that the verification process of ACE-V was seriously flawed and did not constitute peer review under Daubert. We share the judge's consternation with the current verification process. Nevertheless, she found this factor to favor admission, though only slightly, because "limited" review exists on the reliability of ACE-V in forensic publications and because the SWGFAST guidelines outlining ACE-V underwent peer review. With respect to its application to simultaneous impressions, however, the Commonwealth has not introduced evidence of any scientific-technical publication discussing its reliability.

Page 652

Further, the record contains no evidence that SWGFAST, IAI, or any other forensic identification society has promulgated peer-reviewed standards relating to the application of ACE-V to simultaneous impressions. This factor thus also favors exclusion.

iv. Known or potential error rate. We do not quarrel with the motion judge's conclusion that the ACE-V method of fingerprint individualization has a low error rate when used to match a latent fingerprint to a fully inked print of the same finger. We agree that the concern is solely with the rate of false positives. See United States v. Mitchell, 365 F.3d 215, 239 (3d Cir. 2004) ("rate of false negatives is immaterial to the Daubert admissibility of latent fingerprint identification offered to prove positive identification"). The FBI survey and its study of matches between 50,000 latent and full prints, although not without flaws, provide adequate support for the judge's conclusion.

However, the Commonwealth has produced no evidence establishing a similarly low error rate when ACE-V is applied to simultaneous impressions. Neither the FBI survey nor the study involved simultaneous impressions. The record contains no studies regarding the ability of a fingerprint examiner to use simultaneous impressions to effectuate a positive identification and we have not been made aware of any. [Note 19] We recognize, as the motion judge explained, that "the absence of a specific study . . . does not preclude a finding of admissibility." Nonetheless, the absence of any experimentation here, without any other evidence of a low error rate, does not help the Commonwealth. See Canavan's Case, 432

Mass. 304 , 315 (2000) ("We cannot conclude that the . . . mere assertion that a methodology is reliable is sufficient to pass the Lanigan test absent any other evidence showing its reliability"). In the absence of evidence pertaining to the error rate, we conclude that this factor, at best, does not affect our ultimate decision concerning admissibility.

Page 653

v. Standards controlling the technique. The judge concluded that there was no scientific basis for requiring a minimum number of matching points for an individualization. We agree. We are not concerned that this leaves the "evaluation" stage of ACE-V open to the subjective determinations of a fingerprint examiner. A wide variety of experts whose testimony is generally admitted at trial use their education, training, and knowledge to opine on matters about which there does not exist an objective standard. In such instances, "the expert is operating within a vocational framework that may have numerous objective components, but the expert's ultimate opining is likely to depend in some measure on experiential factors that transcend precise measurement and quantification." United States v. Llera Plaza, 188 F. Supp. 2d 549, 571 (E.D. Pa. 2002).

The degree of subjectivity in a fingerprint examiner's ultimate conclusion that a latent print matches a fully inked print seems "of a substantially more restricted compass" than, say, "an electrical engineer's testimony that fire in a clothes [dryer] was caused by a thermostat malfunction." United States v. Llera Plaza, supra at 570, citing Maryland Cas. Co. v. Therm-O- Disc, 137 F.3d 780 (4th Cir. 1998). An examiner follows an objective method laid out in guidelines and standards adopted by SWGFAST. Additionally, the one-discrepancy rule (while not perfect) provides an objective benchmark for examiners. As the motion judge explained, whether a discrepancy is explainable or unexplainable depends on six factors, most of which are objective. [Note 20] The manner in which an examiner's opinion is guided by objective factors makes this process acceptable.

It appears, however, that a fingerprint examiner's opinion regarding the individualization of simultaneous impressions is less bounded by objective factors. Most importantly, although Agent Meagher testified that the ACE-V process does not vary when applied to simultaneous impressions, the record does not establish that either SWGFAST or the IAI has adopted formal

Page 654

guidelines regarding the individualization of simultaneous impressions. There is no standard procedure in place to which an examiner must conform his methods.

The judge also found that the "rigorous qualifications and training requirements for FBI fingerprint examiners . . . help control operation of the ACE-V methodology." Common sense dictates that higher academic and professional standards increase the chances that an expert will properly follow the objective criteria and properly employ his subjective consideration to the facts at hand. This consideration, however, is irrelevant here, where the Commonwealth does not propose to call an FBI examiner as its expert. The Commonwealth's proposed expert is a State trooper, and the original fingerprint examiner was a member of the now disbanded Boston police fingerprint unit. No showing has been made as to the qualifications required for employment and retention at either of these law enforcement agencies. We shall not simply assume that the requirements or expert's qualifications are as substantial as those of the FBI or its fingerprint examiners. See United States v. Llera Plaza, supra at 566 ("Whatever may be the case for other law enforcement agencies, the standards prescribed for qualification as an FBI fingerprint examiner are clear . . . . The uniformity and rigor of these FBI requirements provide substantial assurance that, with respect to certified FBI fingerprint examiners, properly controlling qualification standards are in place and are in force").

Moreover, the Commonwealth does not contend that FBI examiners have confirmed the result of the State examination. To the contrary, in response to a request for confirmation, the FBI issued a report indicating that the simultaneous impressions at issue here were not "of value," apparently concluding that a positive identification could not properly be made using those impressions.

In these circumstances, we conclude that the lack of accepted explicit universal standards controlling the application of ACE-V to simultaneous impressions counsels against admission of this evidence.

3. Conclusion. Evidence of fingerprint individualization determined by application of the ACE-V method to single latent

Page 655

fingerprint impressions meets the Lanigan-Daubert reliability standard. The general acceptance of this application of ACE-V by the fingerprint examiner community leads us to this conclusion. However, the application of ACE-V to simultaneous impressions cannot rely on the more usual application of ACE-V for its admissibility, but must be independently tested against the Lanigan-Daubert standard. On the record before the motion judge, the Commonwealth has not yet established that the application of the ACE-V method to simultaneous impressions is generally accepted by the fingerprint examiner community or that a review of the other Daubert factors favors admission of evidence based on such an application. Consequently, we vacate the judge's supplemental order and remand the case for further proceedings consistent with this opinion.

So ordered.

FOOTNOTES

[Note 1] Patterson was also convicted of armed robbery and two counts of possession of a dangerous weapon.

[Note 2] ACE-V stands for "analysis, comparison, evaluation, and verification." It is the standard methodology used in the United States and many other parts of the world. See infra at part 1.b.

[Note 3] We acknowledge amicus briefs filed by Mark Acree, Robert Bradley, Simon A. Cole, David L. Faigman, Stephen E. Fienberg, Paul C. Giannelli, Lyn Haber, Ralph N. Haber, Donald Kennedy, Jennifer L. Mnookin, Joelle Anne Moreno, Jane C. Moriarty, D. Michael Risinger, John R. Vokey, Sandy L. Zabell, and The New England Innocence Project; National Association of Criminal Defense Lawyers, Massachusetts Association of Criminal Defense Lawyers, and the Committee for Public Counsel Services; the district attorneys for the Berkshire, Cape and Islands, Norfolk, northern, northwestern, and Plymouth districts; and the Secretary of Public Safety.

[Note 4] Although the term ACE-V was not coined until at least 1995, when the Scientific Working Group on Friction Ridge Analysis, Study, and Technology (SWGFAST) documented standards for comparing prints, the steps performed under ACE-V are essentially the same steps performed by fingerprint experts over the last hundred years.

[Note 5] Similarly, Great Britain no longer requires a specific number of Galton points for an examiner to declare a match. For many years, England had required sixteen Galton point matches to make a positive identification. See United States v. Llera Plaza, 188 F. Supp. 2d 549, 555, 567 (ED. Pa. 2002). In 1999, however, the British Court of Appeal (Criminal Division) concluded that fewer than sixteen matching points were needed. See id. at 566, quoting Regina v. Buckley, 143 Si LB 159 (1999) ("If there are fewer than eight similar ridge characteristics, it is highly unlikely that a judge will exercise his discretion to admit such evidence . . . . If there are eight or more similar ridge characteristics, a judge may or may not exercise his or her discretion in favour of admitting the evidence"). According to the Llera Plaza court, the British Court of Appeal explained that a national consensus had developed "that considerably fewer than 16 ridge characteristics would establish a match beyond any doubt." Id. at 567. Additionally, the Court of Appeal had forecast that any type of numerical requirement might be done away with in the near future. The British court cited a 1988 study, commissioned by the Home Office and the Association of Chief Police Officers (ACPO), which concluded "that there was no scientific, logical or statistical basis for the retention of any numerical standard." Id. at 567-568. In 1994, based partially on this study, the ACPO issued a report recommending a completely nonnumerical approach to fingerprint identification. After a fingerprint evidence project board studied the issue in anticipation of a new nationwide system, it recommended the change be made. See id. at 568. In 2001, two years after the Buckley decision, the new nonnumerical system was adopted. The Buckley decision indicates that the nationwide adoption of this plan obviates a bright-line judicial requirement that a positive identification use a specific number of similarities. See United

States v. Llera Plaza, supra, quoting Regina v. Buckley, supra ("If and when [the project board plan is adopted], it may be that fingerprint experts will be able to give their opinions unfettered by any arbitrary numerical thresholds").

[Note 6] Because the Boston police fingerprint unit's latent print section has been suspended from operation, the Commonwealth proposed to offer almost the identical evidence but this time through Detective Lieutenant Kenneth Martin of the State police.

[Note 7] The Commonwealth initially argued that the Daubert-Lanigan hearing should be limited to determining the reliability of the application of ACE-V to simultaneous impressions. The Commonwealth asserted that the general reliability of latent fingerprint identification and ACE-V could be established without recourse to a hearing. The Commonwealth altered that position in early 2004 after it came to light that a man had been wrongfully convicted of armed assault with intent to murder based largely on an erroneous "match" of his fingerprint to a latent print at the crime scene. See Commonwealth v. Cowans, 52 Mass. App. Ct. 811 (2001); Man Freed in 1997 Shooting of Officer, Boston Globe, Jan. 24, 2004, at Al.

[Note 8] The judge used the terms "fingerprint community," "community of fingerprint examiners," and "forensic identification community" interchangeably. Other courts have identified the relevant community in some form of one or more of such terms and we perceive no distinction between these characterizations of the group at issue. See United States v. Mitchell, 365 F.3d 215, 236, 241 (3d Cir. 2004) ("fingerprint examiner community" and "forensic identification community"); United States v. Llera Plaza, 188 F. Supp. 2d 549, 551-552, 563 (E.D. Pa. 2002) ("fingerprint examiner community" and "fingerprint community"); United States v. Sullivan, 246 F. Supp. 2d 700, 703 (E.D. Ky. 2003) ("fingerprint analysis and forensic science fields").

[Note 9] The Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) was established in 1995. Sponsored by the FBI laboratory, the working group includes forty fingerprint experts from various Federal, State, and local agencies throughout North America. Its mission is to formalize and document guidelines and standards that are generally accepted and applied by fingerprint examiners. Its committees develop guidelines and standards, which are subject to critique and debate by all SWGFAST members and, after publication in the Journal of Forensic Identification and presentation at the International Association for Identification, by all members of the fingerprint examiner community. After modification of its guidelines based on this review, SWGFAST republishes them as formal standards. SWGFAST has adopted ACE-V as the standard by which to examine latent fingerprints.

[Note 10] To be admissible, testimony must be relevant as well as reliable. The relevance of identification evidence such as fingerprint analysis is clear and unquestioned by the parties. We thus concentrate on the reliability prong. Ac-cord United States v. Mitchell, 365 F.3d 215, 235 (3d Cir. 2003) ("the fit inquiry in the case of fingerprint identification is not a significant factor, because identity evidence is the archetypal relevant evidence in criminal cases").

[Note 11] Although Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993) (Daubert), itself spoke in terms of scientific knowledge, the Supreme Court, in Kumho Tire Co. v. Carmichael, 526 U.S. 137, 145, 157 (1999), recognized that the Daubert standard was equally applicable to expert testimony based on technical or other specialized knowledge. We adopted this same standard in Canavan's Case, 432 Mass. 304 , 313-314 (2000).

[Note 12] The fingerprint examiner community consists primarily of fingerprint examiners from local, State, Federal, and foreign law enforcement agencies as well as independent or retired examiners. Some of these examiners, such as David Ashbaugh, may spend a significant portion of their time writing, lecturing, and teaching. See, e.g., United States v. Crisp, 324 F.3d 261, 268-269 (4th Cir.), cert. denied, 540 U.S. 888 (2003) (indicating that fingerprint examiners themselves are expert community that suffices for Daubert purposes). Also included are scientists from other fields, such as Dr. Babler, who study the underlying premises of fingerprint examination. The fingerprint community has formed a number of associations and professional groups better to share information and experience, and better to control the standards of their profession. In addition to SWGFAST, many examiners belong to the International Association for Identification (IAI). Founded in 1915, IAI has over 5,000 members. The IAI has established several fingerprint examiner certification programs, publishes the peer-reviewed Journal of Forensic Identification, and awards grants to promote the advancement of forensic science as a profession.

[Note 13] Although the Commonwealth suggests that Commonwealth v. LeClaire, 28 Mass. App. Ct. 932 (1990), accepted evidence of simultaneous impressions, that case is clearly distinguishable. In that case, one of the simultaneous impressions, the thumbprint, was clear and could, on its own, be matched to the defendant. Id. at 933-934.

[Note 14] Similarly, our second suggestion in Commonwealth v. Gaynor, 443 Mass. 245 , 266, 267 (2005), that attacks on the reliability of the specific testing at issue should go to the weight of the evidence followed our analysis of a partially distinct application of the recommended DNA test used by a DNA processor (use of a smaller sample size than set by the manufacturer's test kit) and findings that "Cellmark had conducted validation studies that supported the reliability of testing based on amounts smaller than recommended by the manufacturers" and that the distinct application "has done all that is reason-ably possible to eliminate [the] potential" for distortions.

[Note 15] We cannot surmise the limiting principle by which the Commonwealth's argument would lose force in a case where a fingerprint examiner applied ACE-V to impressions that he believed were left simultaneously and represented two fingers on each hand and two toes on each foot.

[Note 16] Although Ashbaugh did not testify at the hearing below, the Commonwealth offered his 1999 testimony at the Daubert hearing in the Mitchell case. See United States v. Mitchell, 365 F.3d 215 (3d Cir. 2004).

[Note 17] Although in the course of this appeal we have been made aware of an article on simultaneous impressions that allegedly bolsters Meagher's assertion, see Ostrowski, Simultaneous Impressions: Revisiting the Controversy, The Detail (Nov. 05, 2001), an article not in evidence before the judge, the article merely confirms our view that application of the ACE-V methodology is not yet generally accepted in the fingerprint examiner community. In contrast to the FBI survey regarding fingerprint identification generally, the article explains that the author conducted a survey on simultaneous impressions that received only eighteen responses from local, State, and Federal latent print examiners in thirteen States and the District of Columbia. In comparison to the one hundred per cent acceptance of ACE-V methodology in the FBI survey, Ostrowski's survey makes clear that just over fifty per cent of those surveyed would use two or more simultaneous impressions that cannot be identified on their own as the basis for a positive identification. Approximately forty-four per cent of those asked reported requiring at least one of the latent prints to be individually matched in cases of simultaneous impressions and one responding agency requires that each print impression must stand alone. Particularly in light of the extremely small sample in Ostrowski's survey, this hardly amounts to general acceptance in the relevant community.

[Note 18] While David Ashbaugh has proposed several "objective" criteria to use to determine the simultaneity of latent impressions, see Ashbaugh, Quantitative-Qualitative Friction Ridge Analysis 134-135 (1999), it is unclear whether the determination of simultaneity in this case was made using the test Ashbaugh suggests.

[Note 19] The only information provided to us on this narrow issue comes from a postargument letter, which describes a recent National Institute of Science and Technology study that recommends computer identifications only be made by independent individualization of separate fingerprints of simultaneous impressions. While not the basis for our decision, this information adds to our concern that the application of ACE-V in the case at bar may be prone to far more error than the normal use of ACE-V.

[Note 20] These factors are: the condition of the actual friction ridges, the deposition pressure (how hard the finger was placed on the object), the lateral pressure (if the finger was moved after being placed on the object), the texture of the substrate being touched, the method used to process the fingerprint, and the method used to preserve the image.

How to Read a Legal Opinion

A Guide for New Law Students

Orin S. Kerr

This essay is designed to help new law students prepare for the first few weeks of class. It explains what judicial opinions are, how they are structured, and what law students should look for when reading them.

I. What's in a Legal Opinion?

When two people disagree and that disagreement leads to a lawsuit, the lawsuit will sometimes end with a ruling by a judge in favor of one side. The judge will explain the ruling in a written document referred to as an "opinion." The opinion explains what the case is about, discusses the relevant legal principles, and then applies the law to the facts to reach a ruling in favor of one side and against the other.

Modern judicial opinions reflect hundreds of years of history and practice. They usually follow a simple and predictable formula. This

---

Orin Kerr is a professor of law at the George Washington University Law School.

section takes you through the basic formula. It starts with the introductory materials at the top of an opinion and then moves on to the body of the opinion.

The Caption

The first part of the case is the title of the case, known as the "caption." Examples include *Brown v. Board of Education* and *Miranda v. Arizona*. The caption usually tells you the last names of the person who brought the lawsuit and the person who is being sued. These two sides are often referred to as the "parties" or as the "litigants" in the case. For example, if Ms. Smith sues Mr. Jones, the case caption may be *Smith v. Jones* (or, depending on the court, *Jones v. Smith*).

In criminal law, cases are brought by government prosecutors on behalf of the government itself. This means that the government is the named party. For example, if the federal government charges John Doe with a crime, the case caption will be *United States v. Doe*. If a state brings the charges instead, the caption will be *State v. Doe*, *People v. Doe*, or *Commonwealth v. Doe*, depending on the practices of that state.[1]

The Case Citation

Below the case name you will find some letters and numbers. These letters and numbers are the legal citation for the case. A citation tells you the name of the court that decided the case, the law book in which the opinion was published, and the year in which the court decided the case. For example, "U.S. Supreme Court, 485 U.S. 759 (1988)" refers to a U.S. Supreme Court case decided in 1988 that appears in Volume 485 of the *United States Reports* starting at page 759.

The Author of the Opinion

The next information is the name of the judge who wrote the opinion. Most opinions assigned in law school were issued by courts

---

[1] English criminal cases normally will be *Rex v. Doe* or *Regina v. Doe*. Rex and Regina aren't the victims: the words are Latin for "King" and "Queen." During the reign of a King, English courts use "Rex"; during the reign of a Queen, they switch to "Regina."

with multiple judges. The name tells you which judge wrote that particular opinion. In older cases, the opinion often simply states a last name followed by the initial "J." No, judges don't all have the first initial "J." The letter stands for "Judge" or "Justice," depending on the court. On occasion, the opinion will use the Latin phrase "per curiam" instead of a judge's name. Per curiam means "by the court." It signals that the opinion reflects a common view among all the judges rather than the writings of a specific judge.

The Facts of the Case

Now let's move on to the opinion itself. The first part of the body of the opinion presents the facts of the case. In other words, what happened? The facts might be that Andy pulled out a gun and shot Bob. Or maybe Fred agreed to give Sally $100 and then changed his mind. Surprisingly, there are no particular rules for what facts a judge must include in the fact section of an opinion. Sometimes the fact sections are long, and sometimes they are short. Sometimes they are clear and accurate, and other times they are vague or incomplete.

Most discussions of the facts also cover the "procedural history" of the case. The procedural history explains how the legal dispute worked its way through the legal system to the court that is issuing the opinion. It will include various motions, hearings, and trials that occurred after the case was initially filed. Your civil procedure class is all about that kind of stuff; you should pay very close attention to the procedural history of cases when you read assignments for your civil procedure class. The procedural history of cases usually will be less important when you read a case for your other classes.

The Law of the Case

After the opinion presents the facts, it will then discuss the law. Many opinions present the law in two stages. The first stage discusses the general principles of law that are relevant to cases such as the one the court is deciding. This section might explore the history of a particular field of law or may include a discussion of past cases (known as "precedents") that are related to the case the court is deciding. This part of the opinion gives the reader background to help understand the context and significance of the court's decision. The second stage of the legal section applies the general legal principles to the particular facts of the dispute. As you might guess, this part is in many ways the heart of the opinion: It gets to the bottom line of why the court is ruling for one side and against the other.

Concurring and/or Dissenting Opinions

Most of the opinions you read as a law student are "majority" opinions. When a group of judges get together to decide a case, they vote on which side should win and also try to agree on a legal rationale to explain why that side has won. A majority opinion is an opinion joined by the majority of judges on that court. Although most decisions are unanimous, some cases are not. Some judges may disagree and will write a separate opinion offering a different approach. Those opinions are called "concurring opinions" or "dissenting opinions," and they appear after the majority opinion. A "concurring opinion" (sometimes just called a "concurrence") explains a vote in favor of the winning side but based on a different legal rationale. A "dissenting opinion" (sometimes just called a "dissent") explains a vote in favor of the losing side.

II. COMMON LEGAL TERMS FOUND IN OPINIONS

Now that you know what's in a legal opinion, it's time to learn some of the common words you'll find inside them. But first a history lesson, for reasons that should be clear in a minute.

In 1066, William the Conqueror came across the English Channel from what is now France and conquered the land that is today called England. The conquering Normans spoke French and the defeated Saxons spoke Old English. The Normans took over the court system, and their language became the language of the law. For several centuries after the French-speaking Normans took over England, lawyers and judges in English courts spoke in French. When English courts eventually returned to using English, they continued to use many French words.

Why should you care about this ancient history? The American colonists considered themselves Englishmen, so they used the English legal system and adopted its language. This means that American legal opinions today are littered with weird French terms. Examples include plaintiff, defendant, tort, contract, crime, judge, attorney, counsel, court, verdict, party, appeal, evidence, and jury. These words are the everyday language of the American legal system. And they're all from the French, brought to you by William the Conqueror in 1066.

This means that when you read a legal opinion, you'll come across a lot of foreign-sounding words to describe the court system. You need to learn all of these words eventually; you should read cases with a legal dictionary nearby and should look up every word you don't know. But this section will give you a head start by introducing you to some of the most common words, many of which (but not all) are French in origin.

Types of Disputes and the Names of Participants

There are two basic kinds of legal disputes: civil and criminal. In a civil case, one person files a lawsuit against another asking the court to order the other side to pay him money or to do or stop doing something. An award of money is called "damages" and an order to do something or to refrain from doing something is called an "injunction." The person bringing the lawsuit is known as the "plaintiff" and the person sued is called the "defendant."

In criminal cases, there is no plaintiff and no lawsuit. The role of a plaintiff is occupied by a government prosecutor. Instead of filing a lawsuit (or equivalently, "suing" someone), the prosecutor files criminal "charges." Instead of asking for damages or an injunction, the prosecutor asks the court to punish the individual through either jail time or a fine. The government prosecutor is often referred to as "the state," "the prosecution," or simply "the government." The person charged is called the defendant, just like the person sued in a civil case.

In legal disputes, each party ordinarily is represented by a lawyer. Legal opinions use several different words for lawyers, including "attorney" and "counsel." There are some historical differences among these terms, but for the last century or so they have all meant the same thing. When a lawyer addresses a judge in court, she will always address the judge as "your honor," just like lawyers do in the movies. In legal opinions, however, judges will usually refer to themselves as "the Court."

Terms in Appellate Litigation

Most opinions that you read in law school are appellate opinions, which means that they decide the outcome of appeals. An "appeal" is a legal proceeding that considers whether another court's legal decision was right or wrong. After a court has ruled for one side, the losing side may seek review of that decision by filing an appeal before a higher court. The original court is usually known as the trial court, because that's where the trial occurs if there is one. The higher court is known as the appellate or appeals court, as it is the court that hears the appeal.

A single judge presides over trial court proceedings, but appellate cases are decided by panels of several judges. For example, in the federal court system, run by the United States government, a single trial judge known as a District Court judge oversees the trial stage. Cases can be appealed to the next higher court, the Court of Appeals, where cases are decided by panels of three judges known as Circuit Court judges. A side that loses before the Circuit Court can seek review of that decision at the United States Supreme Court. Supreme Court cases are decided by all nine judges. Supreme Court judges are called Justices instead of judges; there is one "Chief Justice" and the other eight are just plain "Justices" (technically they are "Associate Justices," but everyone just calls them "Justices").

During the proceedings before the higher court, the party that lost at the original court and is therefore filing the appeal is usually known as the "appellant." The party that won in the lower court and must defend the lower court's decision is known as the "appellee" (accent on the last syllable). Some older opinions may refer to the appellant as the "plaintiff in error" and the appellee as the "defendant in error." Finally, some courts label an appeal as a "petition," and require the losing party to petition the higher court for relief. In these cases, the party that lost before the lower court and is filing the petition for review is called the "petitioner." The party that won before the lower court and is responding to the petition in the higher court is called the "respondent."

Confused yet? You probably are, but don't worry. You'll read so many cases in the next few weeks that you'll get used to all of this very soon.

III. WHAT YOU NEED TO LEARN FROM READING A CASE

Okay, so you've just read a case for class. You think you understand it, but you're not sure if you learned what your professor wanted you to learn. Here is what professors want students to know after reading a case assigned for class:

Know the Facts

Law professors love the facts. When they call on students in class, they typically begin by asking students to state the facts of a particular case. Facts are important because law is often highly fact-sensitive, which is a fancy way of saying that the proper legal outcome depends on the exact details of what happened. If you don't know the facts, you can't really understand the case and can't understand the law.

Most law students don't appreciate the importance of the facts when they read a case. Students think, "I'm in law school, not fact school; I want to know what the law is, not just what happened in this one case." But trust me: the facts are really important.[2]

[2] If you don't believe me, you should take a look at a few law school exams. It turns out that the most common form of law school exam question presents a long description of a very particular set of facts. It then asks the student to "spot" and analyze the legal issues presented by those facts. These exam questions are known as "issue-spotters," as they test the student's ability to understand the facts and spot the legal issues they raise. As you might imagine, doing well on an issue-

Know the Specific Legal Arguments Made by the Parties

Lawsuits are disputes, and judges only issue opinions when two parties to a dispute disagree on a particular legal question. This means that legal opinions focus on resolving the parties' very specific disagreement. The lawyers, not the judges, take the lead role in framing the issues raised by a case.

In an appeal, for example, the lawyer for the appellant will articulate specific ways in which the lower court was wrong. The appellate court will then look at those arguments and either agree or disagree. (Now you can understand why people pay big bucks for top lawyers; the best lawyers are highly skilled at identifying and articulating their arguments to the court.) Because the lawyers take the lead role in framing the issues, you need to understand exactly what arguments the two sides were making.

Know the Disposition

The "disposition" of a case is the action the court took. It is often announced at the very end of the opinion. For example, an appeals court might "affirm" a lower court decision, upholding it, or it might "reverse" the decision, ruling for the other side. Alternatively, an appeals court might "vacate" the lower court decision, wiping the lower-court decision off the books, and then "remand" the case, sending it back to the lower court for further proceedings. For now, you should keep in mind that when a higher court "affirms" it means that the lower court had it right (in result, if not in reasoning). Words like "reverse," "remand," and "vacate" means that the higher court though the lower court had it wrong.

Understand the Reasoning of the Majority Opinion

To understand the reasoning of an opinion, you should first identify the source of the law the judge applied. Some opinions interpret the Constitution, the founding charter of the government. Other cases

spotter requires developing a careful and nuanced understanding of the importance of the facts. The best way to prepare for that is to read the fact sections of your cases very carefully.

interpret "statutes," which is a fancy name for written laws passed by legislative bodies such as Congress. Still other cases interpret "the common law," which is a term that usually refers to the body of prior case decisions that derive ultimately from pre-1776 English law that the Colonists brought over from England.[3]

In your first year, the opinions that you read in your Torts, Contracts, and Property classes will mostly interpret the common law. Opinions in Criminal Law mostly interpret either the common law or statutes. Finally, opinions in your Civil Procedure casebook will mostly interpret statutory law or the Constitution. The source of law is very important because American law follows a clear hierarchy. Constitutional rules trump statutory (statute-based) rules, and statutory rules trump common law rules.

After you have identified the source of law, you should next identify the method of reasoning that the court used to justify its decision. When a case is governed by a statute, for example, the court usually will simply follow what the statute says. The court's role is narrow in such settings because the legislature has settled the law. Similarly, when past courts have already answered similar questions before, a court may conclude that it is required to reach a particular result because it is bound by the past precedents. This is an application of the judicial practice of "stare decisis," an abbreviation of a Latin phrase meaning "That which has been already decided should remain settled."

In other settings, courts may justify their decisions on public policy grounds. That is, they may pick the rule that they think is the best rule, and they may explain in the opinion why they think that rule is best. This is particularly likely in common law cases where judges are not bound by a statute or constitutional rule. Other courts will rely on morality, fairness, or notions of justice to justify

---

[3] The phrase "common law" started being used about a thousand years ago to refer to laws that were common to all English citizens. Thus, the word "common" in the phrase "common law" means common in the sense of "shared by all," not common in the sense of "not very special." The "common law" was announced in judicial opinions. As a result, you will sometimes hear the phrase "common law" used to refer to areas of judge-made law as opposed to legislatively-made law.

their decisions. Many courts will mix and match, relying on several or even all of these justifications.

Understand the Significance of the Majority Opinion

Some opinions resolve the parties' legal dispute by announcing and applying a clear rule of law that is new to that particular case. That rule is known as the "holding" of the case. Holdings are often contrasted with "dicta" found in an opinion. Dicta refers to legal statements in the opinion not needed to resolve the dispute of the parties; the word is a pluralized abbreviation of the Latin phrase "obiter dictum," which means "a remark by the way."

When a court announces a clear holding, you should take a minute to think about how the court's rule would apply in other situations. During class, professors like to pose "hypotheticals," new sets of facts that are different from those found in the cases you have read. They do this for two reasons. First, it's hard to understand the significance of a legal rule unless you think about how it might apply to lots of different situations. A rule might look good in one setting, but another set of facts might reveal a major problem or ambiguity. Second, judges often reason by "analogy," which means a new case may be governed by an older case when the facts of the new case are similar to those of the older one. This raises the question, which are the legally relevant facts for this particular rule? The best way to evaluate this is to consider new sets of facts. You'll spend a lot of time doing this in class, and you can get a head start on your class discussions by asking the hypotheticals on your own before class begins.

Finally, you should accept that some opinions are vague. Sometimes a court won't explain its reasoning very well, and that forces us to try to figure out what the opinion means. You'll look for the holding of the case but become frustrated because you can't find one. It's not your fault; some opinions are written in a narrow way so that there is no clear holding, and others are just poorly reasoned or written. Rather than trying to fill in the ambiguity with false certainty, try embracing the ambiguity instead. One of the skills of top-flight lawyers is that they know what they don't know: they know when the law is unclear. Indeed, this skill of identifying when a problem is easy and when it is hard (in the sense of being unsettled or unresolved by the courts) is one of the keys to doing very well in law school. The best law students are the ones who recognize and identify these unsettled issues without pretending that they are easy.

Understand Any Concurring and/or Dissenting Opinions

You probably won't believe me at first, but concurrences and dissents are very important. You need to read them carefully. To understand why, you need to appreciate that law is man-made, and Anglo-American law has often been judge-made. Learning to "think like a lawyer" often means learning to think like a judge, which means learning how to evaluate which rules and explanations are strong and which are weak. Courts occasionally say things that are silly, wrongheaded, or confused, and you need to think independently about what judges say.

Concurring and dissenting opinions often do this work for you. Casebook authors edit out any unimportant concurrences and dissents to keep the opinions short. When concurrences and dissents appear in a casebook, it signals that they offer some valuable insights and raise important arguments. Disagreement between the majority opinion and concurring or dissenting opinions often frames the key issue raised by the case; to understand the case, you need to understand the arguments offered in concurring and dissenting opinions.

IV. Why Do Law Professors Use the Case Method?

I'll conclude by stepping back and explaining why law professors bother with the case method. Every law student quickly realizes that law school classes are very different from college classes. Your college professors probably stood at the podium and droned on while you sat back in your chair, safe in your cocoon. You're now starting law school, and it's very different. You're reading about actual cases, real-life disputes, and you're trying to learn about the law by picking up bits and pieces of it from what the opinions tell you. Even weirder, your professors are asking you questions about those opinions, getting everyone to join in a discussion about them. Why the difference?, you may be wondering. Why do law schools use the case method at all?

I think there are two major reasons, one historical and the other practical.

The Historical Reason

The legal system that we have inherited from England is largely judge-focused. The judges have made the law what it is through their written opinions. To understand that law, we need to study the actual decisions that the judges have written. Further, we need to learn to look at law the way that judges look at law. In our system of government, judges can only announce the law when deciding real disputes: they can't just have a press conference and announce a set of legal rules. (This is sometimes referred to as the "case or controversy" requirement; a court has no power to decide an issue unless it is presented by an actual case or controversy before the court.) To look at the law the way that judges do, we need to study actual cases and controversies, just like the judges. In short, we study real cases and disputes because real cases and disputes historically have been the primary source of law.

The Practical Reason

A second reason professors use the case method is that it teaches an essential skill for practicing lawyers. Lawyers represent clients, and clients will want to know how laws apply to them. To advise a client, a lawyer needs to understand exactly how an abstract rule of law will apply to the very specific situations a client might encounter. This is more difficult than you might think, in part because a legal rule that sounds definite and clear in the abstract may prove murky in application. (For example, imagine you go to a public park and see a sign that says "No vehicles in the park." That plainly forbids an automobile, but what about bicycles, wheelchairs, toy automobiles? What about airplanes? Ambulances? Are these "vehicles" for the purpose of the rule or not?) As a result, good lawyers need a vivid imagination; they need to imagine how rules might apply, where they might be unclear, and where they might lead to unexpected outcomes. The case method and the frequent use of hypotheticals will help train your brain to think this way. Learning the law in light of concrete situations will help you deal with particular facts you'll encounter as a practicing lawyer.

Good luck!

PROFFER

The magazine and .357 cartridges were then photographed and referred to the analyst for further review and comparison of the latent prints.

D. Mr. Scott J. Ford

1. Rule 16(a)(1)(G) Summaries: The United States may offer expert testimony on the following subjects:

a. Fingerprint Analysis. The United States will offer expert testimony regarding the fingerprint analysis of items seized in this case on December 31, 2007, specifically, the .357 caliber cartridges. This testimony will be given by Scott J. Ford, Forensic Scientist, of the Hennepin County Sheriff's Office.

2. Expert Qualifications: The United States has identified Mr. Ford as an expert witness in this case in the area of fingerprint analysis. He is qualified as follows:

a. In General: Mr. Ford has been a forensic scientist since 2001, has been employed with the Hennepin County Sheriff's Office since 2008, and presently holds the position of Forensic Scientist, Latent Print Examiner. He holds a Bachelor's degree in chemistry from St. Cloud State University. Mr. Ford has worked in the area for forensic Science, specifically dealing with latent prints, since 2001 where he was employed by the Minnesota BCA. He has received training in the analysis of latent prints and has conducted training in the area of latent print identification.

b. Opinion: Mr. Ford analyzed latent prints found on the

.357 rounds of ammunition. Based on this analysis, Mr. Ford concluded that the latent prints on this ammunition were those the defendant. The latent print found on the magazine was not fully analyzed.

c. Basis for Opinion: Mr. Ford used the process of ACE-V to analyze the latent prints in this case. This is a comparison, evaluation, and verification process. The analysis was conducted by looking at the fingerprint to determine what was present to warrant a comparison. The comparison is conducted by looking at the latent print, determine what areas to compare and target, find a target group, and then look at the minutia for what to target in the known print. The evaluation is a side by side comparison of the latent with the known print. Verification occurs when the identifications are reviewed by another trained examiner.

In this case, Mr. Ford analyzed the latent prints on the .357 cartridges. Based on this review there was no major smudging or distortion. The latents were clean in the levels of detail for ridge flow for pattern, direction, orientation, actual finite detail or points, bifurcations, ridge endings, ridge path, sweat pores, and the appearance of ridges that are thick, fat, or hook at the end. All of these areas were observed in the latent prints. There was some mild pressure distortion. Nine of the points were identified for the minutia. This latent, when compared to the known prints of the defendant revealed that the right index fingerprint of the defendant matched the latent prints found on the .357 cartridges. This comparison was completed using a magnifying glass and importing the prints for the side by side comparison.

E. Ms. Rebecca J. Willis

1. Rule 16(a)(1)(G) Summaries: The United States may offer expert testimony on the following subjects:

a. Chemical Analysis. The United States will offer expert testimony regarding the chemical analysis of controlled substances seized in this case, specifically on December 31, 2007. This testimony will be given by Rebecca J. Willis, Forensic Scientist, of the Minnesota Bureau of Criminal Apprehension

2. Expert Qualifications: The United States has identified Ms. Willis as an expert witness in this case, in the areas of chemical analysis/controlled substance identification. She is qualified as follows:

a. In General: Ms. Willis has been a forensic scientist with the Minnesota Bureau of Criminal Apprehension since April 2007, and presently holds the position of Forensic Scientist. She holds a Bachelor's degree in chemistry from the University of Wisconsin-Madison. She has been employed in various capacities as a chemist since 1998. She also has in-service training in drug analysis, FTIR instrument training, GC/MS instrument training, and clandestine lab certification.

b. Opinion: Ms. Willis conducted chemical analysis on a

CLASS 3

PRACTITIONER GUIDE NO 1

Fundamentals of Probability and Statistical Evidence in Criminal Proceedings

Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses

By

Colin Aitken, Professor of Forensic Statistics, University of Edinburgh,

Paul Roberts, Professor of Criminal Jurisprudence, University of Nottingham

Graham Jackson, Professor of Forensic Science, Abertay University

Prepared under the auspices of the

Royal Statistical Society's *Working Group on Statistics and the Law*

(Chairman: Colin Aitken)

Contents

1. Probability and Statistics in Forensic Contexts

1.1 ***Probability and Statistics – Defined and Distinguished***

Probability and statistics are overlapping but conceptually quite distinct ideas with their own protocols, applications and associated practices. Before proceeding any further it is vital to define these key terms, and to clarify the relationships between them.

Most of this report is devoted to analysing aspects of probability, more particularly to forensic applications of probabilistic inference and probabilistic reasoning. At root, probability is simply one somewhat specialised facet of logical reasoning. It will facilitate comprehension to begin with more commonplace ideas of statistics and statistical evidence.

1.2 Statistics are concerned with the collection and summary of empirical data. Such data are of many different kinds. They may be counts of relevant events or characteristics, such as the number of people who voted Conservative at the last election, or the number of drivers with points on their licenses, or the number of pet owners who said that their cat preferred a particular brand of tinned cat food. Statistical information is utilised in diverse contexts and with a range of applications. Economic data are presented as statistics by the Consumer Price Index. In the medical context there are statistics on such matters as the efficacy of new drugs or treatments, whilst debates on education policy regularly invoke statistics on examination pass rates and comparative levels of literacy.

Statistics may also relate to measurements of various kinds. Familiar examples in criminal proceedings include analyses of the chemical composition of suspicious substances (like drugs or poisons) and measurements of the elemental composition of glass fragments. Whilst these sorts of forensic statistics are routinely incorporated into evidence adduced in criminal trials, any kind of statistical information could in principle become the subject of a contested issue in criminal litigation. These measurements are sometimes known generically as ‘variables’, as they vary from item to item (e.g. variable chemical content of narcotic tablets, variable elemental composition of glass fragments, etc.).

1.3 Probability is a branch of mathematics which aims to conceptualise uncertainty and render it tractable to decision-making. Hence, the field of probability may be thought of as one significant branch of the broader topic of "reasoning under uncertainty".

Assessments of probability depend on two factors: the event *E* whose probability is being considered and the information *I* available to the assessor when the probability of *E* is being considered. The result of such an assessment is *the probability that E occurs, given that I is known*. All probabilities are conditional on particular information. The event *E* can be a disputed event in the past (e.g. whether Crippen killed his wife; whether Shakespeare wrote all the plays conventionally attributed to him) or some future eventuality (e.g. that this ticket will win the National Lottery; that certain individuals will die young, or commit a crime).

The best measure of uncertainty is probability, which measures uncertainty on a scale from 0 to 1. In useful symbolic shorthand, x denotes 'some variable of interest' (it could be an event, outcome, characteristic, or whatever), and $p(x)$ represents 'the probability of x'. An event which is certain to happen (or certainly did happen) is conventionally ascribed a probability of one, thus $p(x) = 1$. An event which is impossible – is certain *not* to happen or have happened – has a probability of zero, $p(x) = 0$. These are, respectively, the upper and lower mathematical limits of probability, and values in between one and zero represent the degree of belief or uncertainty associated with a particular designated event or other variable. Alternatively, probability can be expressed as a percentage, measured on a scale from 0% to 100%. The two scales are equivalent. Given a value on one scale there is one and only one corresponding value on the other scale. Multiplication by 100 takes one from the (0*;* 1) scale to the (0%*;* 100%) scale; division by 100 converts back from the (0%*;* 100%) scale to the (0*;* 1) scale.

Probability can be "objective" (a logical measure of chance, where everyone would be expected to agree to the value of the relevant probability) or "subjective", in the sense that it measures the strength of a person's belief in a particular proposition. Subjective probabilities as measures of belief are exemplified by probabilities associated with sporting events, such as the probability that Red Rum will win the Grand National or the probability that England will win the football World Cup. Legal proceedings rarely need to address objective probabilities (although they are not entirely without forensic applications).[5] The type of probability that arises in criminal proceedings is overwhelmingly of the subjective variety, and this will be the principal focus of these *Practitioner Guides*.

Whether objective expressions of chance or subjective measures of belief, probabilistic calculations of (un)certainty obey the axiomatic laws of probability, the most simple of which is that the full range of probabilities relating to a particular universe of events, etc. must add up to one. For example, the probability that one of the runners will win the Grand National equals one (or very close to one; there is an exceedingly remote chance that none of the runners will finish the race). In the criminal justice context, the accused is either factually guilty or factually innocent: there is no third option. Hence, p(Guilty, G) + p(Innocent, I) = 1. Applying the ordinary rules of number, this further implies that p(G) = 1-p(I); and p(I) = 1-p(G). Note that we are here specifically considering *factual* guilt and innocence, which should not be confused with the *legal* verdicts pronounced by criminal courts, i.e. "guilty" or "not guilty" (or, in Scotland, "not proven"). Investigating the complex relationship between factual guilt and innocence and criminal trial verdicts is beyond the scope of this *Guide*, but suffice it to say that an accused should not be held legally guilty unless he or she is also factually guilty.

Mathematical probabilities obeying these axioms are powerful intellectual tools with important forensic applications. The most significant of these applications are explored and explained in this series of *Practitioner Guides*.

1.4 The inferential logic of probability runs in precisely the opposite direction to the inferential logic of statistics. Statistics are obtained by employing empirical methods to investigate the world, whereas probability is a form of theoretical knowledge that we can project onto the world of experience and events. Probability posits theoretical generalizations (hypotheses) against which empirical experience may be investigated and assessed.

---

[5] Eggleston (1983: 9) mentions the example of proceedings brought under the Betting and Gaming Act 1960, where the fairness of the odds being offered in particular games of chance was in issue.

Consider an unbiased coin, with an equal probability of producing a 'head' or a 'tail' on each coin-toss. This probability is 1 in 2, which is conventionally written as a fraction (1/2) or decimal, 0.5. Using "p" to denote "probability" as before, we can say that, for an unbiased coin, p(head) = p(tail) = 0.5. Probability theory enables us to calculate the probability of any designated event of interest, such as the probability of obtaining three heads in a row, or the probability of obtaining only one tail in five tosses, or the probability that twenty tosses will produce fourteen heads and six tails, etc.

Statistics, by contrast, summarise observed events from which further conclusions about causal processes might be inferred. Suppose we observe a coin tossed twenty times which produces fourteen heads and six tails. How suggestive is that outcome of a biased coin? Intuitively, the result is hardly astonishing for an unbiased coin. In fact, switching back from statistics to probability, it is possible to calculate that fourteen heads or more would be expected to occur about once in every 17 sequences of tossing a fair coin twenty times, albeit that probability theory predicts that the most likely outcome would be ten heads and ten tails if the coin is unbiased. But what if the coin failed to produce any tails in a hundred, or a thousand, or a hundred thousand tosses? At some point in the unbroken sequence of heads we would be prepared to infer the conclusion that the coin, or something else about the coin-tossing experiment, is biased in favour of heads.

1.5 In summary, probabilistic reasoning is logically ***deductive***. It argues from general assumptions and predicates (such as the hypothesis that "this is a fair coin") to particular outcomes (predicted numbers of heads and tails in a sequence of coin-tosses). Statistical reasoning is ***inductive***. It argues from empirical particulars (an observed sequence of coin-tosses) to generalisations about the empirical world (this coin is fair – or, as the case may be, biased). To reiterate: probability projects itself out onto the empirical world; statistics are derived and extracted from it.

1.6 ***Presenting Statistics***

Statistics that summarise data are often represented graphically, using histograms, bar charts, pie charts, or plotted as curves on graphs. Data comprising reported measurements of some relevant characteristic, such as the refractive index of glass fragments, are also often summarised by a single number, which is used to give a rough indication of the size of the measurements recorded.

1.7 The most familiar of these single number summaries is the ***mean*** or average of the data. For the five data-points (counts, measurements, or whatever) 1, 3, 5, 6, 7, for example, the average or mean is their sum (1+3+5+6+7) divided by the number of data-points, in this case 5. In other words, 22 divided by 5, which equals 4.5.

An alternative single number summary is the ***median***, which is the value dividing an ordered data-set into two equal halves; there are as many numbers with values below the median as above it. In the sequence of numbers 1, 3, 5, 6, 7, the median is 5. For an even number of data points, the median is half-way between the two middle values. Thus for the six numbers 1, 3, 5, 6, 7, 8, the median is 5.5. The mean and median are sometimes known as measures of location or central values.

A third way of summarising data in a single number is the ***mode***. The mode is the value which appears most often in a data-set. One might say that the mode is the most popular number. Thus, for the sequence 3, 3, 3, 5, 9, 9, 10, the mode is 3. However, the median of this sequence is 5, and the mean is 6. This simple illustration contains an important and powerful lesson. Equally valid ways of summarizing the same data-set can produce completely different results. The reason is that they highlight different aspects of the data.

1.8 All of these summaries are estimates of the corresponding characteristics (mean, median or mode) of the population from which the sample was taken. In order to assess the quality of an estimate of a population mean it is necessary to consider the extent of variability in the observations in the sample. Not all observations are the same value (people are different heights, for example). What are known as ***measures of dispersion*** consider the spread of data around a central value. One such measure which is frequently encountered in statistical analysis is the ***standard deviation***. The standard deviation is routinely employed in statistical inference to help quantify the level of confidence in the estimation of a population mean (i.e. the mean value in some population of interest). It is calculated by taking the square root of the division of the sum of squared differences between the data and their mean by the sample size minus one. Large values for the standard deviation are associated with highly variable or imprecise data whereas small values correspond to data with little variability or to precise data. At the limit, if all observations are equal (e.g. every observation is 2), their mean will be equal to each observation (the mean of any sequence of observed 2s is 2). By extrapolation, the differences between each observation and the mean will be zero in every case and the standard deviation will be zero.

To illustrate: consider the sample (set of numbers) 1, 3, 5, 7, 9. The sample size is 5 (there are five members of the sample) and the mean is 5 (1+3+5+7+9 =25; 25/5 = 5). The standard deviation is calculated as the square root of

$$[\{(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2\}, \text{divided by } 4]$$

which is the square root of

$$(16 + 4 + 0 + 4 + 16)/4 = 40/4 = 10.$$

The square root of 10 is 3.16, which is the standard deviation for this sample set.

By way of contrast, compare the sample (set of numbers) 3, 4, 5, 6, 7. This sample likewise has five members and a mean of 5. However, the standard deviation is much smaller. It is the square root of

$$[\{(3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2\}, \text{divided by } 4]$$

which is the square root of

$$(4 + 1 + 0 + 1 + 4)/4 = 10/4 = 2.5.$$

Thus, the standard deviation is the square root of 2.5 = 1.58. The smaller value for the standard deviation of the second set of numbers reflects reduced variability (illustrated by the reduced range within which the numbers all fall) in comparison to the first sample set.

1.9 *Statistical Method – Sampling and Confidence Levels*

Statistics relate to a designated "population" of relevant events, individuals, characteristics or measurements, etc. Data collection and analysis encompassing every member of a population of interest (an entire set or "***census***") need not involve probabilistic reasoning at all. However, statistics derived from a sample of a larger population can support inferences about the general population only on the basis of probabilistic reasoning.

Suppose that we wish to survey judicial attitudes regarding the reforms of English hearsay law introduced by the Criminal Justice Act 2003. The relevant population is therefore serving judges in England and Wales. Ideally, we might canvass the attitudes of every single judge through a well-designed questionnaire or interview schedule. Having conducted this research project we might discover, say, that overall 73% of judges are in favour of the reforms, but that 80% think they are too complex whilst 14% believe that we would have been better off leaving the old common law unreformed. There is nothing probabilistic about these statistics, because every member of the relevant population was included in the survey (and by redefining "relevant population", probabilistic calculations could still be avoided without conducting a comprehensive census, e.g. "25% of the twenty judges we interviewed thought that…").

More typically, it is impractical to interview every member of a relevant population and insufficiently rigorous simply to interview an arbitrary subset without any consideration of the methodological implications. Resort to some kind of sampling process is consequently almost inevitable.

1.10 Ideally, a good sample is constituted by a "***random sample***" of the ***target population***, i.e. that group of individuals about whom information is sought. In a random sample, every member of the target population has an equal probability of being selected as part of the sample. One must ensure that the population from which the sample is taken (the ***sampled population***) actually is the target population. Imagine an opinion survey for which the target population is all undergraduates at a particular university. Neither a sample of those students arriving at the university library when it opens on a Monday morning, nor a sample of those students propping up the Union bar at 10.00 p.m. on a Saturday night would successfully match the sampled population to the target population. Sometimes a target population may usefully be divided into sections known as ***strata*** defined by relevant characteristics of interest (in a survey to determine whether the population supports a new law concerning sex discrimination one might wish to stratify by gender to ensure that the views of men and women are represented in proportion to their fractions of the population). A ***stratified sample*** contains suitable proportions from each pertinent stratum of the target population.

In practical contexts, including forensic science and criminal litigation, it is often impossible to identify existing random samples or to generate new ones, stratified or otherwise. Instead, resort must be had to ***convenience samples***, that is, samples conveniently to hand. Diamond (2000) calls these data sets "***nonprobability convenience samples***", underlining their acknowledged lack of randomness. Convenience samples might be, for example, "all glass fragments examined in this particular laboratory over the last five years" or "every shoe-mark comparison that I have seen in my career". The methodological robustness of convenience samples and the legitimacy of their forensic applications are perennially debated. Evett and Weir (1998, 45) comment that "every case must be treated according to the circumstances within which it has occurred, and… it is always a matter of judgement…. In the last analysis, the scientist must also convince a court of the reasonableness of his or her inference within the circumstances as they are presented as evidence".

1.11 One form of inference from sample data to a general population is known as ***estimation***. For example, we might seek to estimate the proportion of all judges in favour of the CJA 2003's hearsay reforms by interviewing a sample of judges. The reliability of any such estimate depends on the appropriateness and robustness of the sampling method employed. A carefully constructed random sampling of, say, 10% of all trial judges in every Crown Court is likely to produce more reliable data – i.e., is likely to be more representative of the population as a whole – than taking a straw poll of the first three judges one happens to encounter in the precincts of the Royal Courts of Justice.

1.12 Statisticians employ probabilistic formulae to measure levels of uncertainty associated with particular estimates. Uncertainty is often expressed in terms of "confidence levels". If a sampling procedure produces a particular statistic – e.g. that 75% of judges polled on balance support the CJA 2003's hearsay reforms – how confident can one be that this result is truly representative of the opinions of the entire population of judges? (Recall that the result of our imaginary census of all judges was a 73% approval rating.) Our random sample might have accidentally included judges with more extreme, or more moderate, opinions than their judicial colleagues. Inclusion of these "outliers" would skew our data – but *ex hypothesi* we do not know whether the 75% statistic derived from our sample over- or under-estimates judicial enthusiasm for the CJA 2003, or is in fact truly representative of the opinions of the entire population of trial judges.

By reference to the size of the sample as a proportion of the entire population of interest (in our example, trial judges in England and Wales) and making certain assumptions about variability in responses, it is possible to calculate confidence intervals for the percentage of CJA-supporting judges across the entire population. We know before conducting any survey that the true percentage of judges who favour the CJA 2003's hearsay reforms must logically lie somewhere between 0% and 100%. We could say that we are 100% confident that the true statistic will lie in this range. As the statistical range narrows, our confidence level will diminish. Taking the 75% judicial approval rating as our datum, we can be more confident that the true figure is within the range 75% plus or minus 10% (i.e., the range 65% - 85%) than in the smaller range 75% plus or minus 2% (i.e., the range 73% - 77%).

1.13 Statisticians routinely combine the ***sample mean*** (the mean value for the sample) with the ***sample standard deviation*** to calculate intervals known as ***confidence intervals*** within which the ***population mean*** (the mean value for the entire population) lies with a certain level of confidence. In this context, "confidence" resembles a probability (although its epistemological status is quite different). Confidence levels are usually expressed as a percentage between 0% and 100%. The wider the interval, the greater confidence one has that the stated confidence interval contains the population mean. Confidence intervals are simply a way of representing uncertainty in estimating the population mean.

The only way to be 100% confident that the interval contains the population mean is to make the interval infinitely wide. This is a logical consequence of uncertainty, which can only be (theoretically) eliminated by including every possible value within the interval. Fortunately, we can construct very short intervals with very high degrees of confidence such as 95% or 99%, which are the "gold standard" in social science research and elsewhere. Results falling outside these confidence levels are declared *statistically significant.*

However, confidence intervals and related judgements of statistical significance are not appropriate measures of the value of evidence in criminal proceedings, for several important reasons. First, the selection of a confidence level is subjective and arbitrary. Why 95%? Why not 99% or 99.9%, or for that matter 75% or 70%? Levels of confidence which are conventionally regarded as satisfactory in social science research have no bearing on the level of confidence ideally required for epistemically warranted verdicts in criminal proceedings. Secondly, employing categorical levels of confidence leads to evidence "falling off a cliff" – i.e., it is excluded entirely - if it falls outside the chosen confidence interval, even by a tiny margin. Evidence which may be highly probative within the stated confidence interval is arbitrarily allotted a value of zero if a small change takes it outside that (arbitrarily chosen) confidence interval. Whatever the merits for social science in proceeding in this fashion, it is plainly unsatisfactory for evidence to be allowed to "fall off a cliff" in criminal proceedings, especially when it is recalled that assessments of statistical significance are merely a way of representing variation in data. Consequently, the fact that a particular estimate falls outside one's preferred confidence interval does not necessarily mean that this result is uninteresting or provides an inaccurate measure of real world events which are themselves subject to natural variation.

1.14 ***Statistical Evidence and Inference***

Statistical inference is the science of interpreting data in order to improve our understanding of events in the world, which in turn may contribute to evidence-based public policy-making. For example, statistical inference from meteorological data might help us to understand climate change and to develop more successful strategies for dealing with it. There is an obvious affinity between statistical inference employing probabilistic reasoning (i.e. reasoning employing probabilities) and criminal adjudication, which is also a form of "reasoning under uncertainty" – we do not know whether the accused is guilty or innocent, and the trial is meant to resolve that issue in a publicly acceptable fashion and to translate it into an appropriate legally-sanctioned verdict.

1.15 It is useful, where possible, to be able to measure uncertainty about issues such as guilt or innocence, so that one can compare levels of uncertainty for different events or different pieces of evidence. One might compare, for example, the probability that the accused is guilty, in light of the evidence adduced at trial – conventionally denoted $p(G|E)$ ("the probability of Guilt, given the Evidence"); and $p(I|E)$, the probability of innocence, given the evidence. These are illustrations of the conditionality of specific probabilities to which reference has already been made. The probability of the event of interest – guilt or innocence – is *conditioned on* (assumes) the evidence adduced at trial. Note the use of the vertical bar | to denote conditioning: to the right of the bar is the assumed known (here E, the evidence); to the left of the bar is the uncertain variable for which a probability is being calculated. In relation to fact-finding in criminal proceedings, this will often be G, guilt; or I, innocence. Since it is certain that the accused is either factually guilty or factually innocent (there is no third option), $p(G|E) + p(I|E) = 1$ (meaning that the probability of Guilt, given the Evidence; plus the probability of Innocence, given the Evidence, logically exhausts the range of all eligible possibilities).

Here, uncertainty is a measure of belief in the truth of the matter at issue (e.g. guilt or innocence). The more strongly it is believed that the accused is guilty, the closer that p(G) will approximate to one. In the criminal justice context, the fact-finder's beliefs are ultimately decisive. Note that, whilst the accused is either factually guilty or not (there is no third option), the measure of one's belief in each of the two possible alternatives can be represented by two probabilities taking any value between zero and one. Where there are two exhaustive and mutually exclusive possibilities, the probability of one can always be calculated if the other is known, e.g. p(G) = 1-p(I); and vice versa, p(I) = 1-p(G). Empirical events are *never* absolutely certain, however, so they can only ever *approximate* one (true) or zero (false). This is just another way of saying that reasoning about empirical events is always, irremediably, reasoning under uncertainty.

1.16 Statistical information may be directly relevant to the matters in issue in criminal proceedings, e.g. in assessing levels of risk involved in particular activities such as driving or operating hazardous machinery. If we wish to know whether the accused was reckless or negligent in causing injury to the victim it is pertinent to know the background or "***base rate***" level of risk for that particular activity. If accidents of a particular sort happen all the time, it is so much less likely that the accused was culpably negligent on this occasion. (Base rates are further discussed in section 2(d), below.)

1.17 Statistics are also a useful way of summarising and presenting pertinent information in legal proceedings. For example, large spreadsheets of data may conveniently be summarised in tables or displayed graphically, and this is entirely appropriate provided that such "demonstrative evidence" is properly understood and that its probative value is competently evaluated.

1.18 As well as contributing items of evidence in the form of statistics, statistical methods can also be employed to interpret data and to evaluate evidence. Examples that might well be encountered in criminal litigation include:

- Reliance on statistical evidence of the quantities of drugs on banknotes, to help the fact-finder to assess – relying on an expert's statistical analysis – whether the banknotes are associated with drug dealing (where the quantities of drugs detected are greater than what might be expected for banknotes in general circulation).

- Reliance on statistical evidence comparing the chemical compositions of drugs from two different seizures, to help the fact-finder to assess – relying on an expert's statistical analysis – whether the seized items originated from the same source.

- Reliance on statistical evidence concerning the occurrence of sudden unexplained infant death (in the general population, or amongst families with particular characteristics), to help the fact-finder to assess – relying on an expert's statistical analysis – whether the occurrence of multiple deaths in any one family should be treated as suspicious.

In each of these illustrations (and countless others that might have been given) statistics are being used, not merely as data with evidential significance for resolving disputed facts which could conceivably be adduced in court as expert evidence, but also as a basis for drawing further inferential conclusions the adequacy of which can be assessed by employing statistical methods and probability theory. Insofar as expert testimony incorporates such statistical or probabilistic reasoning, those experts who produce the evidence, those lawyers who adduce and test it, and those judges and fact-finders who evaluate it all need to grasp the rudiments of statistical inference at a level appropriate to their allotted roles in criminal litigation.

1.19 It is useful to distinguish between two types of sample which typically feature in the evaluation of scientific evidence in criminal proceedings. Unfortunately, there is no standard or agreed terminology to express the relevant distinction, which is between (i) samples of known origin and (ii) samples of unknown origin *relative to an issue in the case*. A sample of unknown origin can be described as the ***recovered sample*** or ***the questioned sample***, whereas samples of known origin are often described as the ***control sample*** or ***reference sample***. The issue is not *where* the sample came from, since samples taken from a crime scene (or victim, or abandoned vehicle, etc.) could be either recovered or control samples, depending on the issue being addressed. The objective is normally to link physical traces associated with an offence to the perpetrator, but sometimes this involves working from samples deposited by an unknown donor at the crime scene or on a victim, etc, and sometimes working in the opposite direction, from samples known to be associated with a suspect or victim which can be linked back to the crime scene or suspect, etc.

For example, fragments of glass collected by an investigator from a broken window at the scene of the crime would be a control sample if the question is: does glass found on the suspect's clothing come from the broken window at the scene of the crime? The origin of the fragments is known to be the window. Similarly, a DNA swab taken from a suspect is a control sample as the origin of the profile is known to be the suspect. Suppose in the first case a suspect is found and fragments of glass are recovered from his clothing. These fragments are a recovered sample, since their origin is unknown: it may or may not be the window at the crime scene. Suppose in the second case a DNA profile is obtained from a blood stain at the crime scene. This is also a recovered sample of unknown origin. It may have come from the suspect, innocently or otherwise, or it may have come from another person entirely.

The control/reference sample may have been taken from a crime scene, victim or suspect. Conversely, a recovered/questioned sample might equally derive from any of these sources. Samples are categorised according to the unknown factor the forensic scientist is seeking to investigate, rather than by reference to their physical location and provenance.

1.20 Finally, statistical methods may be utilised to generate new data with forensic applications (although this may be relatively rare in routine forensic science practice). The first task is to define the forensic problem, which initially confronts investigators and is ultimately determined by jurors in contested criminal trials, e.g., have banknotes recovered from the accused been used in drug dealing activity?[6] A determination is then made as to what information is relevant (e.g. to what extent are banknotes in general circulation contaminated with traces of illegal substances?) and this in turn allows the investigator to assess how a reliable sample might be generated in order to produce new data supporting sound inferential conclusions.

---

[6] Cf. *R* v *Benn and Benn* [2004] EWCA Crim 2100, discussed in §2.22, below.

1.21 We are now beginning to glimpse the power and variety of the potential applications of statistical inference in the administration of criminal justice. It must be stressed, however, that statistical inferences are ultimately only as good as their underlying data, which in turn depends upon (1) the appropriateness of the research design (including sampling methodology) and (2) the integrity of the processes and procedures employed in data collection. Conversely, if data-collection was sloppy and incomplete or samples were poorly chosen, the validity of the inferences drawn from statistical data may be seriously compromised.

1.22 When statistics are being presented and interpreted in forensic contexts (or for that matter, in any other context), there are always two principal dimensions of analysis to be borne in mind:

(i) ***Research methodology and data collection***: Do statistical data faithfully represent and reliably summarise the underlying phenomena of interest? Do they accurately describe relevant features of the empirical world?

(ii) ***The (epistemic) logic of statistical inference***: Do statistical data robustly support the inference(s) which they are said to warrant? Is it appropriate to rely on particular inferential conclusions derived from the data?

2. Basic Concepts of Probabilistic Inference and Evidence

2.1 The first sections of this *Guide* have discussed statistics and statistical evidence in a general way, and introduced some elementary features of probability, including basic notation. In this section and the next we undertake a more systematic and detailed examination of probabilistic reasoning in criminal proceedings.

2.2 The starting point for thinking about information which is statistical or presented in the form of a probability is exactly the same as the starting point for interpreting evidence of *any* kind. The essential issue is: what does the evidence *mean*? The meaning of evidence is a function of the purpose(s) for which it was adduced in the proceedings, which in turn are defined by the issues in the case.

In the context of criminal adjudication, the interpretation of evidence has two principal dimensions. First, the judge must assess whether the evidence is legally admissible. Evidence is admissible if (and only if) it is (i) relevant and (ii) not excluded by an applicable exclusionary rule (such as the hearsay prohibition, rules excluding unfairly prejudicial bad character evidence, or prosecution evidence inconsistent with the demands of a fair trial). Secondly, the fact-finder (jurors or magistrates) must assess the *probative value* of the evidence. This involves determining how the evidence combines with other evidence in the case to support or undermine the prosecution's allegations or the accused's counterclaims. Relevance and probative value are both derived from the logic of inductive inference. Relevant evidence is that which, as a matter of logic and common sense, has some bearing on a fact in issue in the proceedings.[7] The same point can be expressed in terms of probability.[8]

---

[7] "[T]o be relevant the evidence need merely have some tendency in logic and common sense to advance the proposition in issue": *R* v *A* [2002] 1 AC 45, [2001] UKHL 25, [31] *per* Lord Steyn.

[8] Cf. James Fitzjames Stephen, *A Digest of the Law of Evidence* (Stevens, 12th edn, 1948), Art. 1: "any two facts to which [relevance] is applied are so related to each other that according to the common course of events one either taken by itself or in connection with other facts proves or renders probable the past, present or future existence or non-existence of the other".

Evidence is either relevant or irrelevant, legally speaking. There is no middle ground. Probative value (or the "weight" of the evidence) is the measure of the extent to which relevant evidence contributes towards proving, or disproving, a fact in issue. This is a matter of degree.

2.3 Statistical evidence will be relevant and potentially admissible in English criminal proceedings just insofar as it helps to resolve a disputed fact in issue. Probabilistic reasoning will be useful or even indispensable in criminal proceedings if it is needed to interpret statistical evidence or is otherwise a feature of logical inference and common sense reasoning. In order to interpret and evaluate statistical evidence and to assess the adequacy of any probabilistic inferences which the evidence is said to support, criminal justice professionals need to be familiar with a handful of key concepts that statisticians, forensic scientists, and other expert witnesses use to express probabilities and statistical data. These key concepts include:

(a) (absolute and relative) frequencies;
(b) likelihood of the evidence;
(c) the likelihood ratio;
(d) base rates for general issues (prior probabilities);
(e) posterior probabilities;
(f) Bayes' Theorem; and
(g) independence.

This section will explain and illustrate each of these key concepts in turn. It is perhaps worth reiterating that we are not necessarily advocating any of these approaches to conceptualising evidence and inference in criminal adjudication. It is often possible to arrive at particular inferential conclusions simply by applying inductive logic and "common sense" reasoning without needing to resort to mathematical formulations or consciously-articulated probability calculations. Our aim in describing the intellectual tools examined in this section is to make them more readily accessible to readers who might wish to use them and – no less importantly – to help judges, lawyers and forensic scientists monitor, interpret, evaluate and challenge their use by other professionals in the course of criminal proceedings. Section 3 of this *Guide* extends and reinforces the exposition by identifying common errors ("traps for the unwary") and explaining how to avoid them.

2.4 **(a)** ***Frequencies, relative and absolute***

Frequencies are counts of observed events, characteristics or other phenomena of interest to any inquiry. They answer the question: how often does *x* occur? Considered in isolation, such counts produce ***absolute frequencies***. However, it is often more useful to ascertain ***relative frequencies***, that is, frequencies relative to a repeated number of observations (e.g. the frequency of rolling a "6" relative to the number of times a six-sided die is rolled). The relative frequency is the number of occurrences of a feature of interest ("rolling a six"; "drawing an ace from a pack of playing cards"; "finding another person with the same DNA profile"; or whatever), divided by the total number of times the process is repeated.

In the forensic context, stated frequencies normally relate to the occurrence of case-specific evidence, whereas frequencies for the occurrence of issues are usually described as ***base rates***. We will have more to say about base rates in due course (§2.16 – §2.18).

2.5 Frequencies can be illustrated by imagining a roulette wheel with thirty-seven slots, numbered 0-36 in the standard pattern. Consider an experiment (or "trial", in the non-legal sense) in which the wheel is spun 1,000 times and the slot on which the ball lands each time is recorded. The number of times on which the ball lands on a particular slot is the absolute frequency for the number corresponding to that slot. Division of the absolute frequency by 1,000 (the number of spins) gives that slot's *relative* frequency. Similar observed frequencies (absolute and relative) can be recorded for each slot. Relative frequencies are often reported as percentages.

For example, in 1,000 spins the ball might be observed to come to rest in the slot numbered one ("slot no.1") 35 times. This is a relative frequency of 3.5% (35 divided by 1,000). In a fair wheel the ball is equally likely to come to rest in any one of the 37 slots, so the expected number of times the ball would come to rest in slot no.1 is one out of every 37 spins, or 1/37 = 2.7%. Statistical methods can then be used to assess the implications (if any) of this evidence of an observed relative frequency of 3.5% in 1,000 spins against a hypothesis that the wheel is fair with an expected relative frequency of 2.7%. One might want to determine, for example, whether the wheel is fair or biased.

2.6 Assessing the adequacy of an inference is never a purely statistical matter in the final analysis, because the adequacy of an inference is relative to its purpose and what is at stake in any particular context in relying on it. A gambler might treat an observed frequency of 3.5% relative to an expected frequency of 2.7% as sufficient reason for putting his money on no.1, but this discrepancy would not warrant a criminal trial jury drawing the inference that the casino owner is guilty of cheating with a biased roulette wheel. In fact, according to probabilistic calculation, one should expect at least 35 slot no.1s in 1,000 spins about once in every 13 sequences of 1,000 spins. The ultimate inferential conclusion that the evidence proves the accused's guilt beyond reasonable doubt or so that the fact-finder is "sure" is never based solely on the probability of any event; not least because fact-finding in criminal adjudication involves normative issues of juridical classification and moral reasoning (Roberts and Zuckerman, 2010: 133-37). However, the inference of guilt beyond reasonable doubt might well be supported – even very strongly supported – by statistical analysis of relevant data and probabilistic reasoning employing absolute or relative frequencies, where the probability of obtaining particular data (evidence) purely by chance is exceedingly small (unlikely, "beyond reasonable doubt"). Imagine, for example, that the accused claimed to have won the National Lottery jackpot five weeks on the trot or that all three of his bigamous wives on whom he had taken out life insurance accidentally drowned in the bath.[9] At some point in the story, "pure coincidence" as an explanation of apparently incriminating circumstances ceases to retain much plausibility – though it is vital to remember that certain kinds of evidence are prone to replicated error (e.g. a string of eyewitnesses might all misidentify an innocent person as the culprit because she does in fact resemble the real offender).

2.7 Spinning a roulette wheel 1,000 times represents a sample subset of the conceptually infinite population of all possible spins of the wheel. The observed frequency of 3.5% is correspondingly an estimate of the true (relative) frequency of the no.1 slot for that wheel, just as a straw poll of voters attempts to sample the voting intentions of the entire electorate. Successive repetitions of 1,000 spins of the wheel (repeat sampling) would almost certainly produce different estimates of the true frequency. This gives rise to some complex issues of sampling, which are addressed in technical Appendix B.

[9] *R* v *Smith (George Joseph)* (1916) 11 Cr App R 229, CCA.

What is known as the “***error***” in the estimate is a measure of the differences in the estimates produced by repeat sampling, such as repeated experimental trials each comprising 1,000 spins of a roulette wheel. “Error” here is specialist statistical terminology, not to be confused with the commonplace notion of making mistakes. It is not a “mistake” when the roulette wheel produces three slot no.1s on the trot, though this might not be a very good sample from which to generalise because it is so small. For statisticians, “error” models the natural variation in measurements or counts of empirical phenomena, such as spinning a roulette wheel and recording where the ball lands. Error helps to relate the sample to the population. The error can be determined statistically, and this will give us a measure of the quality or “precision” of the estimate. If the precision of the estimates for every slot were calculated over a series of experimental trials, the strength of the evidence supporting the proposition that the wheel is biased could also be calculated. Note that knowledge of the total number of spins (sample size) is essential in order to assess the precision of an estimate. A trial involving 1,000 spins will produce more precise estimates than a trial involving 100 spins, but less precise estimates than a trial involving 10,000 spins of the wheel. Likewise, an inferential conclusion about the fairness or bias of the wheel will be more reliable if it is based on frequencies with calculated measures of precision for all thirty-seven numbers, and not just for the no.1 slot. All else being equal, more data lead to sounder inferences (although no amount of bad data – e.g. those derived from poorly designed experiments or inappropriate samples – will ever reliably warrant inferential conclusions).

2.8 Relative frequencies may in principle be calculated for any population of items, perhaps conceptually unbounded by size. The items might be each individual spin of a roulette wheel or roll of a die, or types of glass, footwear marks, bloodstain patterns, or DNA profiles – relative to, respectively, all spins of the wheel, all rolls of the die, all types of glass seen in a particular laboratory, all types of footwear seen in a laboratory, all bloodstain patterns observed over a period of time, or all DNA profiles in some defined population.

Relative frequencies always state or assume that there is some reference sample against which the frequency of the event in question may be assessed. A further assumption is that this comparison is illuminating and salient for the task in hand. In the context of criminal proceedings, for example, one would expect that a relative frequency would be capable of supporting an intermediate inference about the strength of evidence bearing on disputed facts, leading to the ultimate inference that the accused is innocent or guilty. Nonetheless, there is ample scope for debating the generation of appropriate and meaningful reference samples, and this has occasionally become a bone of contention in criminal appeals.[10]

2.9 Relative frequencies are routinely included in scientific evidence adduced in criminal proceedings. For example, an expert report might contain statements resembling the following:

- "The glass submitted for analysis is seen in approximately 7% of reference glass exhibits examined in this laboratory over the last 5 years."

- "Footwear with the pattern and size of the sole of the defendant's shoe occurred in approximately 2% of burglaries."

- "In one survey of men's clothing, bloodstaining of the quantity and in the pattern seen on the defendant's jacket has been found to occur in 1% of jackets inspected in this laboratory."

It is vital for judges, lawyers and forensic scientists to be able to identify and evaluate the assumptions which lie behind these kinds of statistics. The value of the evidence cannot be ascertained unless its meaning is properly understood. For each of these three examples, the size of the reference sample needs to be considered (the actual number of glass samples examined by the relevant laboratory in the last five years; the number of burglaries from which the 2% statistic was derived; the number of jackets in the survey of men's clothing) .

2.10 One might begin by querying the appropriateness of the reference samples. In relation to the first statement, for example: How were the reference glass exhibits selected? Were

[10] *R* v *Benn and Benn* [2004] EWCA Crim 2100 (employing database of banknotes collected from the Bank of England as a reference sample for banknotes in general circulation); *R* v *Dallagher* [2003] 1 Cr App R 12, CA (earprint expert's database comprised a personal collection of about 600 hundred photographs and 300 earprints).

they just those deriving from criminal investigations? Are glass samples from the catchment area of the laboratory relevant to the current investigation? How, if at all, does the frequency of occurrence of glass types examined in a forensic science laboratory help to evaluate a "match" (whatever that means) between glass fragments found on the clothing of a suspect and glass recovered from a crime scene? Evidence reporting a comparison of the elemental compositions of glass fragments from a crime scene and from a suspect's clothing will be conditioned on various factors, such as the precise location from which fragments were recovered (e.g. the surface of an item of clothing).

Likewise in relation to the second statement, one might ask whether the appropriate reference sample should be limited to footwear marks from burglaries. Do burglars prefer particular kinds of footwear? Do footwear sole patterns differ from year to year? They presumably do for men and women, and between age groups. Might a better reference sample be constructed from sales data from leading retailers? ***Production or sales figures*** data may be adduced in evidence in criminal proceedings as proxies for relative frequency of occurrence , e.g. "Between April 2005 and March 2007, 10,000 pairs of shoes of the same sole pattern and size as the defendant's shoes were sold in 10 outlets in the North of England". This example deliberately highlights many of the assumptions that may be embedded in such data. What is the relevance of the specified dates? Why only in the North of England (do people never travel to buy shoes?) And what percentage of the entire market has been cornered by those 10 outlets? (Is it 10 outlets out of 12, or out of fifty?) The adequacy of a reference sample might be challenged on any or all of these grounds. Unless footwear marks taken from burglaries constitute a perfectly representative sample of footwear ownership amongst the general population (which seems rather unlikely and anyway cannot simply be assumed), choosing an alternative reference sample will produce a different relative frequency. So the construction and selection of reference samples could have a major bearing on the way in which statistical evidence is presented and interpreted. Experts in particular fields may be willing and able to advise on the relative strengths and weaknesses of particular reference samples, or may operate with their own assumptions. Ultimately, however, it is for the legal system to determine whether such data adequately support particular inferences for the purposes of criminal adjudication.

2.11 The selection of items submitted to the laboratory for analysis also involves a sampling process amenable to statistical evaluation. For example, a scene of crime officer (SOCO) or forensic scientist will not submit for scientific analysis every fibre, glass fragment, or blood droplet identified at a crime scene, but will instead make selections of samples to be tested. Any sampling process introduces a risk that the sample will be skewed and unrepresentative, but non-randomized samples of this kind require particularly careful scrutiny. (For obvious reasons, such samples are sometime referred to as ***convenience samples***, but this terminology is not employed consistently and forensic scientists may reserve the term for more systematically collected data as opposed to crime-scene samples). As we have seen, the precision of an estimate can be determined statistically, and may be affected by, amongst other things, the size of the sample. If a desirable level of precision is specified in advance, the sample size can be determined accordingly (e.g. forensic chemists can specify how many tablets of a questioned substance need to be submitted for chemical analysis out of the entire consignment of tablets seized by customs officials), though care must always be taken in specifying the precise nature of the inference drawn from any non-random sample.

.

Notice, again, that statistical reasoning is involved at two discrete stages of this evidential process. First, we can ask how representative of the entire population of items is the sample of items submitted for analysis, e.g. how representative is the sample of glass tested at the laboratory of all the glass pieces that were present at the scene from the broken window? If the answer to this question is or may be "not very", any inferences drawn from the evidence produced by the test will be correspondingly weakened, ultimately to vanishing point. Secondly, assuming that the tested items constitute a representative sample of the glass in the broken window, the evidential significance of finding matching fragments of glass on the suspect's clothing must still be assessed. What is the probative value of this finding, for example, if the matching fragments represent a specified percentage of a designated reference sample, such as "7% of reference glass exhibits examined in this laboratory over the last 5 years"?

2.12 Finally, observe that our illustrative statements employ vague concepts such as "pattern" and "quantity" the meaning of which is not self-evident. When is a series of marks a "pattern"? How precise is the measure of "quantity"? Moreover, what is the relationship (if any) between quantity, pattern and activity, e.g. between blood spatter and violent assault? The value of the evidence adduced in any particular trial cannot be determined satisfactorily unless and until these matters are clarified.

2.13 **(b) *Likelihood of the evidence***

Statisticians and forensic scientists sometimes use the phrase "the likelihood of the evidence". This is shorthand for "the likelihood of finding the evidence in the context of the crime scene and the environment of the suspect" (or its contextual equivalents). References to "likelihood" in this context are often synonyms for "probability". For example, the conclusion that "it is very likely that this correlation would be seen if the suspect were guilty" is equivalent to saying that "there is a high probability that this correlation would be seen if the suspect were guilty".

An expert's assessment of the likelihood (or probability) of obtaining particular findings should be based on data relevant to the type of evidence in question. Relevant "data" are of different types. Towards the harder end of the spectrum, experts may be able to draw on extensive surveys, databases or experimentation. At the softer end of the spectrum, the only available relevant data may be the expert's personal experiences and memories of previous casework.[11] The question is not whether "data" can be assigned to one artificial classification or another – "hard" or "soft" – but rather whether the available data constitute an adequate basis for inferring particular inferential conclusions for particular purposes. Irrespective of their quality and status, data enable the expert to assign a likelihood (or probability) for particular findings that is necessarily personal and subjective, even in relation to ostensibly "hard" data.

2.14 For reasons that will become more apparent as we proceed, it is often illuminating and sometimes essential to express the extent to which evidence supports a particular proposition relative to another proposition in terms of the ratio of two likelihoods: (i) the likelihood of the evidence if one proposition is true; and (ii) the likelihood of the evidence if the other proposition is true. In the context of criminal proceedings, one might compare the likelihood of the evidence, given the prosecution's proposition (e.g. that the accused was at the scene of the crime); as against the likelihood of the evidence, given the defence proposition (e.g. that the accused was not at the scene of the crime).

[11] The Court of Appeal recently endorsed expert witnesses' reliance on personal experiences and unpublished studies in *R* v *Weller* [2010] EWCA Crim 1085.

Suppose that a bloody footwear mark taken from the scene of the crime is said to "match" (in some specified sense of what constitutes a "match") the sole of a shoe in the accused's possession. The probability of finding this evidence of a match if that shoe made the mark (which would be the prosecution's proposition) will often come close to 1 (unless, for example, the shoe has been worn for a considerable time after the commission of the offence, in which case the shoe's tread pattern might have been different at the time of the burglary). Crucially, however, the probability of finding the evidence of a match if *another shoe* made the mark (a possible defence proposition) will be more than 0. For a very rare mark, the probability could be miniscule (approaching zero) but in other circumstances it could be closer to 1, e.g. if the vast majority of the burglars in that area wear the same fashionable training shoes.

These two likelihoods (or probabilities) then represent "the likelihood of the evidence if the prosecution's proposition is true" and "the likelihood of the evidence if the defence proposition is true". The relative values of these two likelihoods provide a measure of the meaning and probative value of the evidence. This is usually represented as a ratio known as ***the likelihood ratio***, which is further elucidated in §2.17, below.

In certain scenarios, the likelihood of the evidence if the defence proposition is true is closely related to the frequency of occurrence of the evidence. For example, if the frequency of occurrence of some characteristic, say males with blue eyes, is estimated at 30% (equivalent to a probability of 0.3) for some specified suspect population, then the probability that a particular male suspect would have blue eyes, *on the assumption that this suspect is actually innocent*, is 30% (or 0.3).

2.15 It is not always possible to obtain a good estimate for a population relative frequency based on sample data: relevant datasets may be incomplete or non-existent. In these circumstances, relative frequencies may be replaced by estimates based on an expert's personal experience and knowledge of the type of evidence in question. Here are some examples:

> "This type of glass occurs in about 10% of the glass samples that I have encountered in the course of my work."

Equivalently:

> "In my experience, one in ten of the glass samples that I have analysed at this laboratory have been glass of this type,"

or

> "From my experience of analysing glass samples at this laboratory, the probability of encountering this type of glass is 0.1."

Observe that the first example expresses the expert's conclusion as a percentage, the second as a proportion (or relative frequency), and the third as a probability (or likelihood). In each case, the progression from data (the expert's personal experience) to inferential conclusion (percentage, proportion or probability) is clearly indicated.

2.16 Whenever such percentages, proportions or probabilities are stated, it is imperative to scrutinise the basis on which the subjective assessment has been made. The person asserting the probability or likelihood should be able to justify it by reference to reasonable assumptions. Probabilities representing subjective measures of belief ideally should be formulated in ways which draw attention to their subjectivity, as the following examples demonstrate (with emphasis):

> "*I estimate* the probability (likelihood) of finding this type and size of shoe sole pattern at scenes of burglary in this area as 2% (or 1/50 or 0.02)."

> "If the defendant had not hit the victim, it is *my opinion* that the probability of finding blood-staining of the quantity and in the pattern seen on his jacket is 1% (or 1/100 or 0.01). I base this estimate on data from a survey of men's clothing."

The second example invites follow-up questioning about the nature of the quoted survey, its sampling and other methodological parameters, and its overall adequacy as a reference sample in relation to the issues in the case. There is an apparent implication that if the defendant *had* hit the victim (the prosecution's proposition) there is a probability higher than 1% (and perhaps substantially higher) of finding this pattern of blood-staining. However, this kind of assertion may express little more than a forensic scientist's intuitive inference from experience. Its underlying assumptions must be identified and opened up to critical scrutiny before one can begin to assess the true value of the evidence in resolving disputed facts.

2.17 **(c) *The likelihood ratio***

As previously stated (and as its name transparently implies), the "likelihood ratio" is an expression of the ratio of two relevant likelihoods (or probabilities). Here is one example (with emphasis) that might be encountered in criminal proceedings:

> "The blood-staining on the jacket of the defendant is approximately *ten times more likely* to be seen if the wearer of the jacket had, rather than had not, hit the victim."

Notice that this likelihood ratio expresses the likelihood of *the evidence*, under the two competing propositions, as opposed to the likelihood of *the act* of hitting. It does *not* state that "given the blood-staining on the jacket, it is ten times more likely that the wearer of the jacket had hit, rather than had not hit, the victim", which is an altogether different proposition introducing many more contingencies than the blood-staining evidence *per se* is capable of addressing.

Our initial example states the value of *the evidence* explicitly conditioned on two competing propositions. This exemplifies the kind of statement that a forensic scientist might write in a report or give in oral testimony. The second, reformulated statement addresses the issue of whether the defendant had or had not hit the victim in the context of the evidence of the blood-staining *and* any other relevant evidence in the case. This is the type of question which fact-finders, rather than expert witnesses, should be left to resolve in contested criminal trials.

2.18 Unfortunately, these two types of statement are frequently confused in practice, producing what is popularly (but not very helpfully) known as "the prosecutor's fallacy". This is one of the principal "traps for the unwary", which is fully explained and, hopefully, neutralised in Section 3 of this *Guide*.

2.19 The likelihood ratio can still be calculated when the evidence is in the form of continuous measurements as opposed to discrete events or characteristics. For example, evidence of the refractive index of glass fragments can be derived from a comparison of two sets of measurements: one set from the control/reference sample (e.g. glass from the scene of the crime) and the other set from the recovered/questioned sample (e.g. glass of unknown origin recovered from the suspect's clothing following his arrest). The value of this evidence can be assessed by considering two competing propositions: (i) that all the glass came from the same source; and (ii) that the recovered sample and the control sample did not come from the same source (i.e. the two samples have different sources). The likelihood ratio of the glass evidence is the ratio of: (i) the likelihood of the observed measurements if the two glass samples share a common source; and (ii) the likelihood of the observed measurements if the two glass samples have different sources.

Forensic scientists and other expert witnesses often translate the numerical likelihood ratio into a verbal formulation expressing a measure of strength for a particular proposition. For example, the expert might state that:

> "My findings provide moderate [weak/strong/very strong/etc] support for the theory that the accused, rather than some other person, was the driver of the car used in the robbery."

Alternatively, some experts employ a numerical scale (e.g. a six- or ten-point scale) as a more jury-friendly proxy for the likelihood ratio or as a more intuitive and looser quantification of the probative value of their evidence.[12] In whatever way the likelihood ratio (or other asserted measure of probative value) is translated into evidence, and even if the likelihood is presented in its raw numerical form, it is essential that advocates, judges and fact-finders are able to interpret its true meaning and thereby assess the probative value of the evidence. Experts themselves can and should provide vital assistance by clearly acknowledging their use of a conventional linguistic or numerical scale to express the strength of evidential support, and explaining how it maps onto the likelihood ratio, in their written statements and testimony.

---

[12] Cf. *R* v *Atkins* [2009] EWCA Crim 1876; *R* v *Shillibier* [2006] EWCA Crim 793; *R* v *Bilal* [2005] EWCA Crim 1555.

2.20 **(d) *Base rates for general variables (prior probabilities)***

Base rates (sometimes also called ***background rates***) are the relative frequencies of variables in a general population before consideration of special circumstances or evidence relating to the case in hand. These do not need to be expressed quantitatively. For example (using fictitious statistics merely for the purposes of illustration):

- "The incidence of death directly attributable to Drug A is 85% of all deaths of abusers of Drug A. "

- "Death attributable to natural hypoglycaemia in elderly non-diabetic patients is extremely rare."

Whereas frequencies relate to specific *evidence*, base rates refer to general *events and other background variables*. This vital distinction further clarifies the respective roles of expert witnesses and fact-finders in criminal proceedings. Base rates for general variables are independent of case-specific information, to which they form the backdrop. Thus, base rates may well be introduced by a competent expert before another expert presents, and takes into account, the results of their own examinations. Base rates can also be used to assign ***prior probabilities*** for those events. The term "prior" encapsulates the fact that such probabilities are developed *prior* to any evidence specific to the instant case.

The first question for the expert (or for the first expert) in each of our examples would be 'what is the base rate for the event or characteristic in question (general prevalence among Drug A abusers of death directly attributable to Drug A; incidence of death caused by natural hypoglycaemia in elderly non-diabetic patients)'? The second question for the expert (or the question for the second expert) is 'what is the probability of their findings, taking account of the prosecution's and defence's competing propositions'? In the first example, the expert would need to consider data on (i) levels of Drug A found in drug-abusers who had died as a consequences of ingesting Drug A; and (ii) levels of Drug A found in drug-abusers who had not died as a direct result of ingesting Drug A. In our second example, the expert would derive a likelihood ratio from data on (i) the levels of insulin found in elderly non-diabetic patients who had died through natural hypoglycaemia; and (ii) data relating to deaths in similar patients resulting from induced hypoglycaemia.

Finally, the question for the fact-finder (*not* the expert) in each case is 'what is the probability, in light of the evidence adduced at trial, that the accused is guilty (that the death was directly attributable to the drug; that the elderly deceased was injected with insulin, etc.)'?

2.21 Base rates can have significant implications for inferential conclusions. Imagine a medical diagnostic test with a high probability of a positive result if the patient has the disease (this measure is known as ***sensitivity***) and a high probability of a negative result if the patient does not have the disease (known as ***specificity***). This hypothetical diagnostic test, then, is both very sensitive to the presence of the disease and very specific to it. Suppose that a particular patient is diagnosed with the disease. What is the probability that the patient actually has the disease? The fact that the diagnostic test is both very sensitive and very specific does not, as might be thought, guarantee that a positive diagnosis is very likely to be correct. This is a function of base-rates. Imagine that nobody in the region has the disease (the base-rate is zero). No matter how sensitive and specific the diagnostic test is – perhaps it only errs one time in a million – on this assumption every single positive diagnosis will be wrong. The probability of a correct diagnosis when the base rate is zero is zero, irrespective of the diagnostic power of the test.

2.22 Base rates that are derived from samples (as distinct from those derived from a census) invite methodological questions paralleling those concerns previously identified in relation to calculating the relative frequencies of evidence. Base rates will be a poor base-line for any inferential purpose if data collection was poorly executed or the sampling procedure was methodologically flawed. Even if base rates supply methodologically robust information for some purposes, they will not necessarily serve to illuminate the matters specifically in issue in criminal proceedings. As in relation to frequencies of evidence, one must carefully scrutinise the inferential link, if any, between background base rates and the issues requiring proof in the current trial.

A different sample drawn from the same population will, almost certainly, give a different answer for a relative frequency. This does not mean that either or both of these frequencies is "wrong". Rather, both frequencies are (different) estimates of a true unknown rate. In *Benn and Benn*,[13] a case concerned with base rates for trace quantities of drugs on banknotes in general circulation, the Court of Appeal remarked that, "the question of the validity of a database depends upon the purpose which is to be served".[14] Deficiencies in the database were not considered fatal to the safety of the convictions where "the comparison made between the notes in the appellants' possession and the database was merely part of the prosecution case showing a connection between the appellants and the cocaine". Whilst the value of the statistical evidence was thus marginalised as merely part of the general background to the prosecution's case, the Court did not really consider that sampling deficiencies arguably robbed this evidence of any discernable meaning or probative value.

As we saw in 1.13, the reliability of an estimate can be determined by specifying, at a stated level of confidence (e.g. 95% or 99%), an interval within which the true rate is thought to fall. The narrower the interval for a given confidence level, the more reliable the estimate.

2.23 **(e) *Posterior probabilities***

All probabilities are predicated (or "conditioned") on specified assumptions. This is merely another way of expressing the inherent conditionality of probability as a species of reasoning under uncertainty. Thus, for example, one might calculate the probability that the accused is guilty, given the evidence that has been presented in the trial – in mathematical notation, $p(G|E)$. Whereas base rates for general variables inform prior probabilities, conditional probabilities conditioned on case-specific events or evidence can be described as ***posterior probabilities*** – such as the probability that the accused is guilty after (posterior to) having heard all the evidence. The ultimate posterior probability, of guilt or innocence and their corresponding legal verdicts, is always a question for the fact-finder in English and Scottish criminal proceedings.

2.24 Expert witnesses must not trespass on the province of the jury by commenting directly on the accused's guilt or innocence, and should generally confine their testimony to

[13] *R* v *Benn and Benn* [2004] EWCA Crim 2100.

[14] ibid. [44].

presenting the likelihood of their evidence under competing propositions. However, experts are not absolutely precluded from stating posterior probabilities relating to intermediate facts proving or constituting the offence, if invited to do so by the court and providing that such statements are appropriately qualified and contextualised. The court must understand, and be prepared to accept, the suppositions on which statements such as the following are predicated:

- "In my opinion, it is highly likely that the defendant kicked the victim."

- "I believe there is a 99% chance (probability of 0.99) that the defendant handled explosives."

- "In my opinion, the accused is very likely to have been the author of the ransom note."

All these statements relate specifically to evidential facts and only indirectly to the ultimate issue of guilt or innocence. It may be helpful, in appropriate cases, for expert witnesses to express their conclusions in this form (also note that our examples commendably flag up the subjective nature of the inference as the expert's "opinion", "belief", etc.). However, it is vital to appreciate that posterior probabilities relate to disputed facts rather than to information adduced in evidence, and the two must never be confused. Experts normally testify to ***relative frequencies*** (to inform likelihoods of the occurrence of evidence), or occasionally to ***base rates*** (***prior probabilities***), rather than to the truth or falsity of contested issues in the trial (***posterior probabilities***). Where experts depart from the norm by testifying directly to posterior probabilities, they should do so deliberately and advisedly, not merely through confusion. Insofar as experts do testify to posterior probabilities, they must spell out and justify the conditioning assumptions and prior probabilities supposedly warranting them.

2.25 **(f)** ***Bayes' Theorem***

Bayes' Theorem is a mathematical formula that can be applied to update probabilities of issues in the light of new evidence. One begins with a prior probability of an issue and some pertinent item of evidence. Bayes' Theorem calculates a posterior probability for the issue, conditioned on the combined value of the prior probability *and* the likelihood ratio for the evidence. This posterior probability can then be treated as a new prior probability to which a further additional piece of evidence can be added, and a new posterior probability calculated (now taking account of the original prior probability and the likelihood ratios for both pieces of evidence). The process can be repeated over and over, finally resulting in a posterior probability conditioned on the entire corpus of evidence in the case.

Fact-finding in criminal adjudication is, generally speaking, accomplished by ordinary common sense reasoning rather than through the application of mathematical formulae, as the Court of Appeal emphatically reiterated in *Adams*.[15] It should be borne in mind, however, that although most evidence adduced in criminal proceedings does not come with a pre-assigned quantified numerical value attached (e.g. what is the probability that an eyewitness identification is accurate? Or the probability that a confession is true?), much forensic science evidence (including DNA profiling) *is* predicated on quantified probabilities and is consequently directly amenable to Bayesian calculations. Moreover, even unquantified evidence can be assigned a subjective probability in Bayesian reasoning. Bayes Theorem is a codification of the reasoning that should be applied in the assessment of evidence. It is a statement of logic. Its application ensures evidence is assessed rationally.

2.26 Bayes' Theorem is best illustrated through a simple artificial example. Consider a population of interest comprising 1,000,001 people. One person has committed a burglary, the other million are innocent. Suppose that by chance 1% of the innocent people (1,000,000/100 = 10,000) have carpet fibres on their clothing matching the carpet at the burgled premises. Assume that the burglar's clothing also picked up these fibres during the burglary. These distributions are summarised in Table 2.1:

[15] *R* v *Adams (No 2)* [1998] 1 Cr App R 377, CA.

Table 2.1: **Numbers of innocent and guilty people on whom fibres are present and absent**

Fibres	Guilty	Innocent	Total
Present	1	10,000	10,001
Absent	0	990,000	990,000
Total	1	1,000,000	1,000,001

We can read off from the final right-hand column of the 'Present' row that the fibres were found on 10,001 individuals – 10,000 of whom are innocent and one of whom is the guilty burglar.

From these data we can construct ***prior probabilities*** for guilt and innocence, before the evidence of the fibres is considered. The prior probability of guilt is 1/1,000,001 – one person out of 1,000,001 is guilty. In other words, the probability that a person selected at random from the population would be guilty is 1/1,000,001. Complementarily, the probability that a person selected at random from the population is innocent is 1,000,000/ 1,000,001.

The ***posterior probabilities*** for guilt and innocence can be obtained from the row labelled "Present" in which there are 10,001 people of whom 1 is guilty. Thus, after the evidence of the fibres is considered, the posterior probability of guilt is 1/10,001 – one person out of 10,001 is guilty. In other words, the probability a person selected at random from the population on whom relevant fibres are found would be guilty is 1/10,001. Complementarily, the probability that a person selected at random from the population on whom relevant fibres are found is innocent is 10,000/10,001.

The ***likelihood ratio*** is the ratio of the probability for the presence of the relevant fibres amongst the guilty (the proportion of people in Table 2.1's Guilty column for whom the fibres are present) to the probability for the presence of the relevant fibres amongst the innocent (the proportion of people in Table 2.1's Innocent column for whom the fibres are present). In this simple example, the probability for the presence of the relevant fibres amongst the guilty is one divided by one, i.e. 1. The probability for the presence of the relevant fibres amongst the innocent is 10,000 divided by 1,000,000, or 1/100. The ratio of these probabilities is 1 divided by 1/100 which is 100. This may be summarised in words as "the evidence of the presence of relevant fibres is one hundred times more likely if the person is guilty than if the person is innocent".

2.27 These probabilities can also be expressed, equivalently, in terms of ***odds ratios***. The prior odds a person selected at random from the population is guilty are given by the ratio of the two prior probabilities for guilt and innocence, namely 1/1,000,001 divided by 1,000,000/1,000,001 or 1 to 1,000,000. This equates to saying that, "the odds are one million to one against guilt for a person selected at random from the population"; or that "the odds are one million to one in favour of innocence for a person selected at random from the population".

The posterior odds that a person selected at random from the population on whom relevant fibres are found is guilty are given by the ratio of the two posterior probabilities, namely 1/10,001 divided by 10,000/10,001 or 1 to 10,000. This equates to saying that, "the odds are ten thousand to one against guilt for a person selected at random from the population on whom relevant fibres are found"; or that "the odds are ten thousand to one in favour of innocence for a person selected at random from the population on whom relevant fibres are found".

2.28 Bayes' Theorem links the prior odds, the posterior odds, and the likelihood ratio in the following way:

$$\textit{posterior odds} = \textit{likelihood ratio} \times \textit{prior odds.}$$

That is to say, the posterior odds are calculated by multiplying together the likelihood ratio and the prior odds (or again, the posterior odds are the product of the likelihood ratio and the prior odds).

In our example, the prior odds are one in a million, the posterior odds are one in ten thousand and the likelihood ratio is one hundred. This is a verification of Bayes' Theorem, since one in ten thousand is 100 times one in a million. Of course, it is not necessary to apply the sledgehammer of Bayes' Theorem to crack this simple example, the results of which could be obtained more or less directly by common sense mathematical calculation.

Bayes' Theorem comes into its own, and may have significant forensic applications, when the calculations are more complex and the issues to be addressed may not be so self-evident.

2.29 Bayes' Theorem can be expressed more formally and in a way which applies directly to criminal proceedings, as follows:

The ***posterior odds*** in favour of the prosecution proposition are equal to the product of:

(i) the ratio of the probability of the evidence if the prosecution's proposition is true, to the probability of the evidence if the defence proposition is true (i.e. ***the likelihood ratio***); and

(ii) the ***prior odds*** in favour of the prosecution proposition.

Referring back to Table 2.1, the evidence is the presence of fibres on the clothing of a suspect (recovered sample) that are of the same type and colour as carpet fibres at the crime scene (control sample). The prosecution proposition is that the suspect is guilty of the crime. The defence proposition is that the suspect is innocent. The likelihood of the evidence given (conditioned on) the truth of the prosecution's proposition is 1, or $p(E|G) = 1$. The likelihood of the evidence given (conditioned on) the truth of the defence proposition is 10,000/1,000,000 = 1/100, or $p(E|I) = 1/100$. The probability of guilt given the evidence – $p(G|E)$ – is 1/10,001. The probability of innocence given the evidence – $p(I|E)$ – is 10,000/10,001.

Notice that the first pair of quantities is conditioned *on the assumption of guilt* or *of innocence* (as the case may be), whereas the second pair of quantities is conditioned *on the evidence*. Moving from the first pair to the second pair of quantities involves ***transposing the conditional***. It can be see that "E", representing the evidence, occupies the position to the left of the conditioning bar in the first pair of quantities, whereas in the second pair its position has shifted (been transposed) to the right of the bar. Bayes' Theorem can be described as a logical and legitimate procedure for transposing the conditional. Illegitimate transposition of the conditional is (for better or worse) widely known as "the prosecutor's fallacy", which is explained and debunked in Section 3 of this *Guide*.

2.30 A second illustration demonstrates the power of Bayes' Theorem as a formula for updating conditional probabilities and should help to clarify its current and potential forensic applications.

Suppose that an accident is caused by an unidentified bus. A total of 1,000 buses are in service in the vicinity. Blue Bus Company owns 90% of these 1,000 buses and Red Bus Company owns the remaining 10%. An eyewitness testifies that the bus that caused the accident was Red. However, a psychologist gives uncontradicted expert testimony that eyewitness identifications of this type are accurate only about 80% of the time. That is to say, an eyewitness will report seeing a Red (or Blue) bus when the bus was truly Red (or Blue) 80% of the time. Conversely, the eyewitness will report that the bus was Red when it was Blue (or Blue when it was Red) 20% of the time.

The entire population of interest comprises 900 Blue buses (90% of 1,000) and 100 Red Buses (10% of 1,000). If the accident was in fact caused by a Blue bus, the eyewitness would accurately report 720 Blue buses (80% of 900) and misidentify the other 180 (20% of 900) as Red. If the accident was in fact caused by a Red bus, the eyewitness would accurately report 80 Red buses (80% of 100) and misidentify the other 20 (20% of 100) as Blue. On this scenario, a Red bus is four times more likely to be reported as Red than Blue. However, *a priori* there are nine times as many Blue buses as Red buses operating in the area. These results are summarised in Table 2.2.

Table 2.2: Numbers of Red and Blue buses, as reported and in fact

	Actually Blue	Actually Red	Total
Reported Blue	720	20	740
Reported Red	180	80	260
Total	900	100	1000

2.31 Bayes' Theorem states that the posterior odds are equal to the likelihood ratio multiplied by the prior odds.

The prior odds are 9:1 (or, simply, 9) in favour of a Blue bus having caused the accident, or p(Blue) = 0.9. Complementarily, p(Red) = 0.1. The prior odds in favour of a Red bus are the reciprocal of the odds in favour of a Blue bus and are hence 1:9 (or 1/9).

We are told that the eyewitness testifies that the bus involved in the accident was Red. The likelihood ratio is the probability that a bus is reported as Red given that it is Red (80/100) divided by the probability that it is reported as Red given that it is Blue (180/900), which equals 4. The posterior odds of the bus being Red when reported Red are the product of the prior odds and the likelihood ratio, 1/9 multiplied by 4 which equals 4/9. The corresponding probability of a bus being Red given it is reported as Red is then 4/13 and the probability of a bus being Blue, given it is reported as Red is the complement of this, namely 9/13 (Check that the ratio of these two probabilities is 4/9 (or 4:9), the odds.)

It seems counterintuitive that the evidence should favour the bus being Blue when the eyewitness testified Red: but Table 2.2 and Bayes' Theorem both corroborate that conclusion. It is obvious at the outset that – all else being equal – a Blue bus was much (nine times) more likely to be involved in the accident than a Red bus. The eyewitness testimony decreases these prior odds to posterior odds of 9:4. Nonetheless, given the eyewitness's stipulated error rate (20%), when the eyewitness testifies Red this actually favours Blue by a ratio of 180:80, or 9:4 – as can be read off from the "Reported Red" row of Table 2.2. Bayes' Theorem powerfully confirms this counter-intuitive result. The likelihood ratio of 4 reduces the odds in favour of Blue from 9:1 to 9:4. In other words, the eyewitness evidence supports the proposition that the bus is Red, but not with sufficient probative force to make it *more likely than not* that the bus is Red, all things considered. This would require a probability greater than 0.5, or (equivalently) odds greater than 1:1 – "a fifty-fifty chance", as we might say. (Note, however, that this conclusion is alarming for real-world litigation only on the supposition that eyewitnesses really do confuse red and blue 20% of the time, and – to our knowledge – there is no empirical evidence warranting that assumption.)

The purpose of the example is two-fold. First, it provides a numerical verification of Bayes Theorem. Second, it shows how consideration of uncertainty about the accuracy of an eyewitness may be included in the evaluation of the evidence of the eyewitness. One can model the effect of various levels of uncertainty on the value of the evidence of the colour of the bus that was involved in the accident.

2.32 **(g) *Independence***

The concept of independence is central to both legal proof and mathematical probability. In law, two or more independent items of evidence may be mutually corroborative. This is first and foremost a logical rule of inference which the law of criminal procedure sometimes elevates into a formal legal requirement (most formal corroboration requirements have been abolished in England and Wales, but Scottish law still retains a general demand for corroboration in serious criminal cases). The logic of corroboration through independent evidence extrapolated to probability theory by ***the product rule for independent events***, which states that, if two events are independent, the probability of both of them occurring together (known as their *conjunction*) can be calculated by multiplying together the probability of the first event and the probability of the second event. These propositions are best demonstrated through simple illustrations using coin-tossing and playing cards.

2.33 Two events are independent in the probabilistic sense if the occurrence of one has no bearing on the probability of the occurrence of the other. Successive outcomes of the tosses of a coin or of tosses of several different coins are independent. Consider two fair coins, which when tossed are (by definition, as "fair" coins) equally likely to produce a head or a tail. The occurrence of a head when the first coin is tossed has no effect on the probability of a head when the second coin is tossed. On the toss of the first coin, the probability of a head is equal to the probability of a tail, which equals ½ or 0.5. These probabilities remain the same for the toss of the second coin and on subsequent tosses of these or of other fair coins. Independence holds no matter how many times the process is repeated.

Consider one fair coin which is tossed twice. The probability of two heads in two tosses of the coin is (utilising the product rule) ½ x ½ = ¼, or p(two heads) = 0.25. The probability of two tails is exactly the same. However, the probability of one head and one tail is ½, or p(head and tail) = 0.5. This is because there are two ways in which the outcome of the two tosses of the coin can be a head and a tail: head followed by tail, or tail followed by head. The probability of each of these two sequences is ¼, and the probability of either one or the other (known as their ***disjunction***) is calculated by adding (*not* multiplying) the probability of each, as we say, *exclusive* event. In other words, ¼ + ¼ = ½. The events (head followed by tail) and (tail followed by head) are known as exclusive events since one occurs to the logical exclusion of the other.

2.34 Now consider a slightly more complicated example. A normal pack of playing cards contains 52 cards in four suits, spades (♠), hearts (♥), diamonds (♦) and clubs (♣) with thirteen cards in each suit. The pack is well-shuffled. A card is picked from the pack at random, i.e. in such a way that each card is equally likely to be selected. Suppose that the ace of spades (*A*♠) is the card drawn at random from the pack. This card is replaced, the pack is well-shuffled and then a second card is drawn, again at random. This process of selection is described as ***selection with replacement***. These successive draws of cards are independent events. Replacing the first card drawn and then shuffling the pack ensures that the outcome of the first draw has no effect on the outcome of the second draw. In other words, the outcomes of the two draws are independent. The probability that the card drawn at the second draw is also the *A*♠ is the same as the probability that the first card was the *A*♠, 1/52. Given that these are (as we have stipulated) independent events, the product rule applies, so that the probability of drawing the *A*♠ twice in succession is 1/52 x 1/52 = 1/2,704.

The same type of calculation can be extended to groups of cards. For example, the probability that a card picked at random from a pack is a ♠ is 13/52 = ¼, or p(spade) = 0.25. There are 13 ♠ in the pack, and each is equally likely to be selected.

2.35 If two or more events are not independent, then they are ***dependent***. There is also a product rule for calculating the probability of the conjunction of dependent events. Consider again the selection of two cards at random from a normal pack, one after the other. This time, the first card selected is *not* returned to the pack after it has been viewed, so that the second card is drawn from a reduced pack of 51 cards. This type of selection process is called ***selection without replacement***.

What is the probability of selecting two aces without replacement? It is the product of the following two probabilities:

(i) the probability that the first card selected is an ace, which is 4/52 = 1/13; *and*

(ii) the probability that the second card selected is also an ace, which is 3/51 = 1/17 (since there are now only 51 cards remaining in the pack, of which 3 are aces).

Thus, the combined probability, p(drawing two aces without replacement) is 1/13 x 1/17 = 1/221.

This result can also be derived and demonstrated by direct enumeration. The order in which the cards are drawn is significant. There are twelve ways of drawing two Aces, viz (♠♥), (♥♠), (♠♦), (♦♠), (♠♣), (♣♠), (♥♦), (♦♥), (♥♣) (♣♥), (♦♣) and (♣♦). There are 52 ways of choosing the first card without replacement and 51 ways of choosing the second card from the reduced deck. There are therefore 52 x 51 = 2,652 equally likely ways of choosing two cards from a pack of 52 cards. Of these 2,652 ways, twelve give two aces. Thus the probability of drawing two aces equals 12/2,652 = 1/221.

2.36 In this example, the probability of the second event (drawing an Ace) is dependent on the first event (drawing an Ace). The probability of drawing a second Ace (i.e., assuming an Ace was drawn on the first draw) is 1/17. This is less than the probability (1/13) of drawing an Ace on the first draw. However, two events may also be associated in such a way as to *increase* the probability of the second event relative to the probability of the first event. This somewhat counterintuitive result is illustrated in Appendix B.

2.37 The probabilistic foundations of games of chance have real-world analogues in criminal litigation. It is therefore vital for criminal practitioners to grasp the fundamentals of probabilistic thinking, and these fundamentals include the concept of independence. The nature of the dependency in examples involving packs of cards or tosses of coins is readily identifiable. In real life the dependencies are typically more difficult to ascertain. Yet as Section 3 will elucidate, it is a serious error to apply the simple product rule to events that are not, or may not be, independent. As a general rule of thumb, independence should be verified and demonstrated and not merely *assumed* by default.

3. Interpreting Probabilistic Evidence – Anticipating Traps for the Unwary

3.1 Reasoning errors in criminal adjudication are by no means confined to information concerned with probabilities. However, probability, statistical evidence, and inferential reasoning associated with them do seem to be especially prone to recurrent errors and misinterpretation. Statistical and probabilistic evidence are typically adduced in court through the medium of a scientific report or expert witness testimony adduced at trial. There is consequently considerable overlap between an examination of probabilistic evidence and reasoning in criminal proceedings and the general topic of expert evidence, as previous sections have already intimated.

3.2 This section begins by drawing attention to some fundamental principles for correctly interpreting reports or testimony provided by forensic scientists and other expert witnesses. We will emphasise, in particular:

(a) the importance of correctly identifying the level of the propositions addressed by the evidence, in order to interpret its real bearing (if any) on the issues in the case; and

(b) the nuanced language used by scientists to express their inferential conclusions, which requires a certain amount of "unpacking" in order to decode its true meaning.

Thereafter, the following analytically distinct (though in practice, often compounded) reasoning errors will be examined and elucidated:

(c) illegitimately transposing the conditional ("the prosecutor's fallacy");

(d) source probability error;

(e) underestimating the value of probabilistic evidence;

(f) probability ("another match") error;

(g) numerical conversion error;

(h) false positive fallacy;

(i) fallacious inferences of uniqueness; and

(j) unwarranted assumptions of independence.

Learning about these reasoning errors as an abstract intellectual exercise is not the same as successfully avoiding them in practice. Their twisted logic can seem enormously seductive and they are frequently perpetrated by professionals who ought to know better, especially in pressured situations such as giving evidence in criminal trials. This is all the more reason for lawyers and judges, as well as forensic scientists and other expert witnesses, to study the recurrent errors in probabilistic reasoning examined in this section. Forewarned is forearmed. Knowing what to look out for, coupled with eternal vigilance, is the best way to guard against falling into traps for the unwary.

3.3 **(a)** ***Relating the evidence to the issue: what question does the expert's evidence purport to answer?***

Expert evidence (or indeed, any other evidence adduced in criminal proceedings) might be conceptualised as offering an answer to a question. The ultimate question in criminal adjudication is always: is the accused guilty or innocent of the offence(s) charged? Of course, in deference to the presumption of innocence the ultimate question in English and Scottish criminal proceedings is not framed in this way. Instead, we ask: has the prosecution proved the accused's guilt beyond reasonable doubt (or so that the fact-finder is "sure" of the accused's guilt)?

Expert evidence does not answer the ultimate question directly; this is a matter solely within the province of the fact-finder. Instead, expert evidence addresses intermediate evidential facts with a bearing on the ultimate issue. For example, an expert might testify that glass found on the accused's clothing resembles (or "matches") glass from the scene of the crime; or that the accused's fingerprints are similar to (or "match")[16] those on the window of the burgled house; or that the type of firearm discharge residue (FDR) evidence found on the victim of a shooting supports the proposition that the accused's gun fired the

[16] The notion of a forensic science expert "declaring a match", though familiar, is problematic. In the first place, the criteria for declaring "a match" may be contested amongst practitioners, or may be eminently contestable even where most or all competent practitioners agree on conventional criteria for determining what constitutes a match. More fundamentally, if all trace evidence ultimately rests on probabilistic calculations, experts perpetrate source probability error (discussed in (d), below) whenever they conclusively assert "a match".

fatal shot. It is then a matter for the fact-finder to determine whether this evidence, taken together with all the other evidence in the case, is sufficient to warrant a finding of guilt.

When one grasps that evidence (including expert evidence) is adduced by the prosecution or defence to answer a particular question, it follows that the meaning and value of that evidence cannot be determined without first identifying the original question. One cannot assess whether evidence is successful in proving a matter in issue until one knows what the issue is and how the evidence relates to it. This observation might sound banal; but it is not. In fact, nearly all of the reasoning errors described in this section are either variations on, or are at least exacerbated by, an elementary failure to identify, with sufficient care and particularity, the question which the evidence is capable of answering.

3.4 A useful starting point in evaluating expert evidence is to identify the ***level of proposition*** (or type of answer) which the evidence addresses. Four different levels of proposition can usefully be distinguished:

(i) source level propositions;
(ii) sub-source level propositions;
(iii) activity level propositions; and
(iv) offence level propositions.

Each of these levels of proposition is regularly encountered in criminal litigation.

3.5 The following are examples of pairs of complementary ***source level propositions***:

- "The defendant is the source of the semen at the crime scene."/
  "The defendant is *not* the source of the semen at the crime scene."

- "The defendant's sweater is the source of the fibres at the crime scene."/
  "The defendant's sweater is *not* the source of the fibres at the crime scene."

- "The damaged window frame is the source of the paint fragments recovered from the defendant's clothing."/
  "The damaged window frame is *not* the source of the paint fragments recovered from the defendant's clothing."

The value of evidence adduced in support of source level propositions is usually related to the relative frequency of the characteristic of interest. Suppose this frequency is one in a thousand (1/1,000). As a first approximation, the value of evidence can be expressed as the reciprocal ("one over") of that relative frequency, e.g. one divided by 1/1,000 or 1,000. For each of our three pairs of example source level propositions, there must be some reference sample (e.g. a database of DNA profiles; records of fibres recovered from crime scenes; or previous analyses of paint fragments found on clothing examined at the laboratory) allowing the expert to calculate the probability of the evidence if it came from an alternative source consistent with the accused's innocence. Notice that source level propositions do not say anything about how the evidence came to be at the scene or on the defendant's clothing, nor do they take into account such variables as the quantity, position or distribution of the recovered material. Source level propositions are limited to addressing whether or not a piece of evidence came from a particular source. Assessment of evidence under source level propositions requires little in the way of circumstantial information.

3.6 Certain forensic science techniques, notably DNA profiling, have become so sensitive that it may be desirable to formulate expert evidence with greater circumspection and precision in terms of ***sub-source level propositions***, such as the following:

- "The DNA recovered from the crime sample came from Mr Smith."
- "The DNA recovered from the crime sample did not come from Mr Smith;" or "The DNA recovered from the crime sample came from some other person."

Sub-source level propositions introduce a greater degree of caution by taking the inferential process, as it were, one stage further back. The expert does not make any direct assertion about the type of biological material from which the DNA was ostensibly extracted (e.g., the semen or blood recovered from the crime scene). Rather, the evidence is restricted to the sub-source or cellular level – leaving open the possibility that the material from which the DNA has been extracted may not be the assumed, asserted or most obvious source. For example, biological samples recovered from the crime scene might contain mixtures of different types of cellular material – saliva, skin cells, secretions, etc. – contributed by several human donors. In these situations, it will be very unlikely that the scientist is able to attribute the DNA to any one type of cellular material.

3.7 Running in the opposite direction, activity level propositions are more coarse-grained and potentially provide more probative evidence than source level propositions. The following are examples of ***activity level propositions***:

- "The defendant had intercourse with the victim."
- "The defendant walked on the carpet in the burgled house."
- "The defendant smashed the window."

The expert is now addressing the issue of whether or not the accused actually did something (had intercourse; walked on a carpet; smashed a window, etc.), not merely whether or not physical evidence might have come from a specified source or sub-source. This is unavoidably controversial territory. In order to arrive at the value of the evidence assuming an activity level proposition, the expert needs to factor into their analysis much more than merely relative frequencies. For example, it may be necessary to consider the physics of transfer and persistence of physical evidence, with associated subjective probabilities. It is also necessary to take into account any innocent explanations offered by the accused for the existence of apparently incriminating evidence. For example, an accused may say that his clothing had been sprayed by the victim's blood when he, an innocent passerby, attempted to render first aid to the dying victim. The scientist's role in this situation is to assess the likelihood of obtaining the pattern and distribution of blood-staining that had been observed on the clothing if the accused's suggestion were, or might have been, true.

Crucially, in terms of the balance and usefulness of scientific findings, consideration of activity level propositions provides an assessment of the probative value of the absence of material ("missing evidence"); something that cannot be assessed if source (or sub-source) propositions are considered.

3.8 ***Offence level propositions*** are the most coarse-grained and probatively consequential of all the types of statement that might be encountered in expert witnesses' reports or testimony. They take the following form:

- "The defendant raped the complainant."
- "The defendant burgled the house."
- "The defendant committed criminal damage."

Offence level statements assert conclusions about criminal responsibility and liability, which are paradigmatically questions for the court. Expert witnesses should not testify to propositions at the offence level, because they involve factual and moral judgments that forensic scientists are not jurisdictionally competent to make (e.g. Did the victim consent? Was harm caused unlawfully?). Of course, it does not necessarily follow that, in practice, forensic scientists and other expert witnesses are always successful in steering clear of offence level propositions, sometimes there is a trespass beyond the logical scope of their evidence.

3.9 *Practitioner Guide No 4* will present a more systematic analysis of interpretational issues relating to the different levels at which evidential propositions may be stated. For these introductory purposes, it will suffice to underline three fundamental points.

First, it is essential on every occasion to identify the precise question which scientific evidence is being adduced to answer. Testimony offered to answer the question, "What is the source (or even sub-source) of this evidence?" is plainly not equivalent to testimony answering the question, "Did the accused have intercourse with the complainant?", still less does it answer the ultimate question, "Did the accused rape the victim?" Note that these are *all* questions for the fact-finder in criminal proceedings, since they all require inferential conclusions to answer them, albeit at different levels of proposition. That testimony or other evidence is being adduced to answer a particular question does *not* entail that the expert witness should try to answer that question *directly*. Generally speaking, expert witnesses should avoid stating inferential conclusions and instead restrict themselves to commenting on the ***likelihood of the evidence*** under each of two competing propositions, i.e. to expressing and explaining the ***likelihood ratio***.

Secondly, there is a delicate balance to be struck between the transparency and scientific rigour of an expert's evidence and its potential helpfulness to the court. Sub-source propositions are the most rigorous and transparent, but they may not go very far in resolving disputed questions of fact and could be open to misinterpretation (e.g. without guidance, the fact-finder could easily mistake a sub-source level proposition for a source level proposition). Source level propositions, likewise, may have limited utility for criminal adjudication. Even if source level testimony substantially warrants a particular inference, e.g. that a suspect is the source of a blood stain, this does not help determine whether the stains were transferred during a criminal assault or entirely innocently or by a third party. Activity-level propositions come closest to the questions that the fact-finder has to answer, but often build in more speculation and assumptions. The scientist may be able to draw on further relevant expertise, e.g. about transfer and persistence for trace evidence,[17] that can be factored into an activity level proposition and provide valuable assistance to the fact-finder. In every case, however, it is essential that everybody in the courtroom understands the significance of what is being said, that the scientist's assumptions and inferential reasoning should be available to critical scrutiny, and that expert witnesses are able to explain and justify the reasonableness of their assumptions if called upon to do so.

Thirdly, it is worth repeating that evidence evaluation is always a fundamentally comparative enterprise. At all levels of proposition the scientist needs to consider the ***likelihood ratio*** for the evidence, i.e. the probability of the evidence given the prosecution proposition, compared with the probability of the evidence given the defence proposition. Ascertaining the prosecution proposition is normally fairly straightforward, e.g. "the accused is the source of the crime stain at the scene" (paving the way to potential further inferences, that the accused was present at the scene, and that he committed the offence there). It may be more difficult to generate realistic defence propositions if there has been limited pre-trial defence disclosure, although it is always possible to use the negation of the prosecution's proposition as a default setting ("the accused did *not* leave the crime stain at the scene", etc.). Postulating appropriate propositions for comparison is closely tied to the facts of each case, and it is a largely intuitive, non-mathematical exercise, rooted in "logic and experience" (in the sense familiar to criminal lawyers). These important issues affecting the value and interpretation of probabilistic evidence will be further explored and elucidated in *Practitioner Guide No 4*.

[17] *R* v *Weller* [2010] EWCA Crim 1085; *R* v *Reed and Reed*; *R* v *Garmson* [2010] 1 Cr App R 23; [2009] EWCA Crim 2698.

3.10 **(b) *Interpreting the language of inferential conclusions***

It is also important to pay close attention to the precise language used in expert reports and testimony to express evidentially significant connections between phenomena (and expert witnesses should correspondingly take care to express such connections precisely). Many forensic scientists and other experts employ stock terminology in report-writing which, although a valid way of expressing preliminary conclusions, may be of limited value to a court and could potentially be misleading unless appropriately qualified and interpreted with circumspection. Further discussion of these ideas may be found in Jackson (2009).

3.11 **"Consistent with"**: It is sometimes said that the evidence is "consistent with" a particular proposition relating to a contested issue in the case, e.g.:

> "Traces of chemicals detected on the swab from the right hand of the suspect *are consistent with* coming from the explosive used at the scene of the explosion."

To say that something is "consistent with" something else means only that the stated proposition (hypothesis) is not excluded by the evidence. It says nothing about how likely the proposition is to be true. For example, buying a ticket is consistent with winning the National Lottery, but it does not make winning very likely. Buying a ticket is also consistent with *not* winning the National Lottery, and this second outcome is very much more likely than the first, though both are equally "consistent with" the premiss (buying a ticket).

3.12 **"Could have come from" / "Could have originated from"**: Once a "match" (however defined) has been obtained between a control sample and a recovered sample, it is common practice for scientists to express an inferential conclusion, such as the following:

- "The semen stain could have come from Mister X, the suspect."
- "The footwear mark at the crime scene could have been made by the shoe the accused was wearing."
- "The blood stain on the window-frame could have been left by the defendant."

- "The fibres recovered from the defendant's clothing could have originated from the victim's sweater."
- "The person shown holding the knife in the CCTV footage could be the defendant."

Statements such as these might be understood as establishing a proven association between the crime and the accused. Notice, however, that "could have come from" does not rule out other possible sources. Indeed, it does not even say that the identified source is the most likely candidate. There may well be other explanations that have not been offered, or even considered, by the scientist, including explanations with a higher probability than the association specified in each statement. Like expressions of "consistency", variations on "could have come from" or "could have originated from" give absolutely no indication of the likelihood that the postulated source is the actual source of the evidence.

3.13 **"Cannot be excluded"**: Another phrase commonly employed by expert witnesses is "cannot be excluded", as in the following examples:

- "The defendant cannot be excluded as the stain donor."
- "The victim cannot be excluded as the source of the blood spatter on the accused's shirt."
- "The broken window cannot be excluded as the source of the glass in the defendant's shoe."

"Cannot be excluded" is the mirror-image of "could have come from" in its vagueness, and is equally susceptible to misinterpretation. There may be any number of alternative sources or explanations that likewise "cannot be excluded", and some of these might be much more likely. The fact that a postulated source cannot be excluded does not mean that evidence of association is strongly or even more than minimally probative.

3.14 A particular variant of the "cannot be excluded" formula is common in DNA and paternity cases, where it is expressed as ***the probability of exclusion***. This probability states what proportion of the population the characteristic would exclude, regardless of who is the donor of the crime-stain. For example, if a relevant characteristic is shared by 0.1% (a relative frequency of 0.001 or 1/1,000) of the population, then the probability of exclusion is 0.999. If a characteristic is present only in 0.1% of the population then it is absent in 99.9% of the population. Thus, if the characteristic is present at the scene of the crime and identified as coming from the (unidentified) perpetrator, 99.9% of the population are excluded as donors of the characteristic.

The probability of exclusion answers the question: "How likely is this characteristic to exclude Mister X if he is *not* the donor of the stain?" However, this could be a very misleading way of expressing the probative value of the evidence, because the court is normally interested in a completely different question: "How much more likely is the evidence if Mister X *is* the donor of the stain than if some randomly selected person were the donor?" (i.e. ***the likelihood ratio***). The probability of exclusion does not address this second, forensically salient question, the answer to which turns crucially on the size of the suspect population. If the relevant population is, say, 1 million, there will be 1,000 individuals with the relevant characteristic, notwithstanding a probability of exclusion of 99.9%.

3.15 Misinterpretations of the probability of exclusion set the pattern for most of the other recalcitrant reasoning errors identified in this section. The trump card, in every case, is scrupulous attention to the meaning of a particular proposition Always ask: what question does this evidence purport to answer? On what assumptions is this statement of probability conditioned? Avoiding elementary probabilistic reasoning errors is as banal and intensely difficult in practice as that.

3.16 **(c)** ***Illegitimately transposing the conditional ("the prosecutor's fallacy")***

Several references have already been made to the probabilistic reasoning error popularly known as "the prosecutor's fallacy", but more technically and accurately described as illegitimately ***transposing the conditional***. This is an error that in principle any participant in criminal proceedings could make: lawyers, judges, jurors, or forensic scientists. In many ways, forensic scientists who fall into this error could be regarded as the chief culprits, since if the expert makes a transpositional error in their initial report or testimony it is eminently foreseeable that lawyers, judges and fact-finders will simply adopt and perpetuate it. After all, the expert is supposed to be the expert and lawyers, judges and lay fact-finders claim no special expertise in reasoning with probabilities. However, erroneous transpositions of the conditional have repeatedly been exposed in scientific evidence – especially DNA profiling testimony – adduced by the prosecution, and illegitimately transposing the conditional has for this reason widely come to be known as "the prosecutor's fallacy". Although not truly apt, the label has stuck.

We saw in Section 2(f) 2.25-2.31, above, that Bayes' Theorem transposes the conditional *legitimately* by employing a valid mathematical formula for this purpose. We are now concerned with evidential propositions which purport to transpose the conditional illegitimately, without employing Bayes' Theorem or any other recognised method of producing a valid conclusion. The error is typically perpetrated unconsciously, and is consequently all the more insidious and liable to precipitate miscarriages of justice for being hidden even from those ostensibly best equipped to avoid it.

3.17 The most direct way of conceptualising the error is to say that it confuses ("transposes") the conditioning event. Consider the following two propositions:

#1: If I am a monkey, I have two arms and two legs.

#2: If I have two arms and two legs, I am a monkey.

These conditional propositions ("if….") are clearly not equivalent![18] Proposition #1 is true, whereas proposition #2 is false. Moreover, proposition #2 patently does not follow from proposition #1. When criminal justice professionals illegitimately transpose the conditional they perpetrate an error equivalent to treating proposition #1 as though it were the equivalent of, or at least an authorised version of, proposition #2.

3.18 Utilising shorthand probabilistic notation, the last example can be expressed as follows:

$p(A+L|M) \neq p(M|A+L)$;

[18] Another example of patently non-transitive conditional propositions: #1 "If I am reading this Guide, I can read English"; #2 "If I can read English, I am reading this Guide".

i.e. the probability of Arms and Legs, given (assuming; conditioned on) Monkey is not equal to the probability of Monkey, given (assuming; conditioned on) Arms and Legs.

In the context of criminal proceedings, the standard form of the error confuses the probability of finding the evidence on an innocent person with the probability that a person on whom the evidence is found is innocent, i.e.

$$p(E|I) \neq p(I|E).$$

Mathematical notation is particularly useful here, because we can see that “E” and “I” have changed places. On the left hand side of the equation, the conditioning event is “I” (“assuming innocence”). On the right hand side of the equation, “I” has swapped places with “E”, which has moved to the left side of the bar indicating the conditioning event (“assuming the evidence”). The conditional has been transposed. These are absolutely not equivalent expressions, as indicated by the “does not equal” sign (≠) dividing the equation.

We have repeatedly stated that the value of evidence is always conditioned on particular assumptions, which should be specified. Consider the following pair of questions about the value of evidence:

> *Assuming that the accused is innocent*, what would be the probability of finding this trace evidence on him?
>
> *Assuming that this trace evidence has been found on the accused*, what is the probability that he is innocent?

The italicised part of each question is the assumption on which the relevant probability is conditioned. The conditional is illegitimately transposed in criminal adjudication when questions of the first type are misrepresented or misinterpreted as questions of the second type.

3.19 A more elaborate illustration should help to make these abstract propositions more readily comprehensible.

Suppose that the DNA profile of a suspect matches the DNA profile from a blood stain found at a crime scene. Assume that the DNA profile has a relative frequency in the *relevant population* of 1/1,000, i.e. one in every thousand people in that country has a matching DNA profile. Let us also stipulate that the relevant *suspect population* (specified through other, non-probabilistic, considerations such as geographical proximity and opportunity) contains 10,001 individuals, the offender and 10,000 innocent others.

One member of the suspect population has been arrested, swabbed, and found to have a DNA profile that matches the profile of the crime stain. Since the relative frequency of the DNA profile in the general population is 1/1,000, the *expected* number of suspects with matching profiles is 10,000 x 1/1,000 = 10. These would be entirely random or "adventitious" matches with entirely innocent individuals. If the offender is known to be the 10,001st member of the suspect population, there are an expected 11 people in the suspect population with matching profiles – ten (expected) "random matches" plus the one offender. It should be emphasised that this "expected" number is a probabilistic projection, not an empirically-observed frequency. Eleven matches are "expected" in exactly the same sense as the "expected" number of heads in ten tosses of a fair coin is five.

3.20 Having been told that the relative frequency of the DNA profile in the general population is 1 in 1,000, it is tempting to equate this to the probability that the suspect is innocent. In other words, to consider the probability of guilt to be 999/1,000; or in notational shorthand, p(Innocent) = 1/1,000; p(Guilty) = 1 – 1/1000 = 999/1,000. But this involves illegitimately transposing the conditional! The stated frequency of 1/1,000 does not represent the probability of the suspect's innocence, but rather the probability that a person picked at random from the general population would have a matching profile, *irrespective of any connection to the offence*.

There are 10,001 people in our suspect population. A particular suspect has been found to have a profile which matches the profile of the crime stain. If the matching profile were the only evidence available, the probability of the suspect's being innocent would be 10/11, which implies p(Guilty) = 1 – 10/11 = 1/11, or 0.09. A probability of 0.09 is not even close to proof on the balance of probabilities, let alone proof beyond reasonable doubt. Yet the error of transposing the conditional produced a fake p(Guilty) of 999/1000 = 0.999, which would easily constitute proof beyond reasonable doubt according to most commentators and participants in empirical research (always allowing for the fact that the courts resolutely refuse to quantify the criminal standard of proof, doubtless for good reason). This stylised illustration demonstrates just how devastatingly powerful such a reasoning error could be in lending credibility to unwarranted conclusions and possibly contributing towards miscarriages of justice.

3.21 Some real-world examples of criminal appeals in which the conditional was illegitimately transposed at trial are given in Appendix B. The so-called "prosecutor's fallacy" tends to be associated with DNA evidence. This is understandable inasmuch as DNA evidence involves quantified probabilities which are articulated in court as random match probabilities, thus routinely presenting opportunities for communication breakdown of one kind or another potentially involving illegitimate transpositions of the conditional.

However, it cannot be stressed too strongly or too often that illegitimate transpositions are *not* a peculiar feature of DNA evidence, but rather potentially could infect every type of evidence, including in particular all kinds of scientific and other expert evidence adduced in criminal proceedings. This follows from the fact that all types of evidence can be assigned subjective probabilities (taking account of relevant data, where available). For example, an expert might testify that there is an 80% probability that mud recovered from the accused's car came from the riverbank near where the deceased's body was discovered;[19] or that there is a "distinct possibility" (perhaps 40%) that handwriting on a forged cheque is the accused's.[20] It would obviously be a crass error to misinterpret these probabilities, respectively, as "an 80% chance of guilt" or "a 40% chance of guilt" of the offences charged. However, both these illustrations of expert testimony involve a more insidious variant of illegitimate transposition, which is described in the next section. The general lesson is that the conditional may be illegitimately transposed *whether or not the evidence is explicitly quantified* and *whether or not expert witnesses realise that they are implicitly drawing upon or assuming probabilistic calculations*.

3.22 **(d) *Source probability error***

[19] *R* v *Shillibier* [2006] EWCA Crim 793, [71].

[20] *R* v *Bilal* [2005] EWCA Crim 1555, [7] – [8].

When illegitimate transpositions of the conditional occur in relation to source level propositions, this is more technically known as *source probability error*.

3.23 Suppose that a crime has been committed, and trace evidence is recovered linking a suspect to the scene, e.g. a DNA match between blood from a murder victim and blood recovered from the suspect's clothing. A scientist determines a value for the frequency of the DNA profile in a relevant population as 1 in 7 million, and writes a report stating:

> "The probability that the blood on the clothing of the suspect came from someone other than the victim is 1 in 7 million. This implies that, with a complementary probability of 6,999,999/ 7 million, the blood on the suspect's clothing came from the victim."

The stated conclusions are unwarranted. They comment erroneously on the source of the blood recovered from the suspect's clothing. It would be legitimate for the scientist to say that, *if* the blood on the clothing of the suspect did not come from the victim, there would be a 1 in 7 million probability of matching the victim's DNA profile. But this is not a proposition about the likelihood of the source; it is the ***random match probability***. In order to calculate the probability that the victim is the source of the blood it would be necessary to know the size of the relevant population (and possibly much else besides, e.g., the probability of an error in testing or of contamination of samples). If there were, say, 14 million potential blood-donors in the relevant population (and making the simplifying assumptions that there is no other pertinent evidence in the case and that all 14 million potential donors were antecedently equally likely to be the true source), the probability that the matching blood came from the victim would be 1/3 (the real victim plus the two other "expected" random matches in the population).

The scientist in this example has transposed the conditional between p(finding a match, assuming the blood on the suspect's clothing could have come from anybody in the relevant population) and p(the blood on the suspect's clothing came from a source other than the victim, assuming a match), i.e. p(Match | Innocent Source) ≠ p(Innocent Source | Match). The scientist then correctly calculates that p(Victim's DNA | Match) = 1 – p(Innocent Source | Match), but irreparable damage has already been done by the initial illegitimate transposition of the conditional. On our assumed frequencies of occurrence in the relevant population, 1 – p(Innocent Source | Match) = 1 – 2/3 = 1/3; again, nowhere near the erroneously asserted value for p(Victim's DNA | Match) of 6,999,999/7 million.

3.24 Returning to the non-DNA illustrations mentioned at the end of the last section (and ignoring for these purposes any complications regarding what constitutes "a match"), the probability that the mud has a common source in the first example is *not* 80%; and the probability that the handwriting in the second example is the accused's is *not* 40%. Rather, these probabilities represent the probability that the recovered sample matches the control sample, *assuming a common source*: p (M | S). To calculate the probability of a common source, p (S | M), it is necessary to factor in the probability that the samples would match, even match perfectly, notwithstanding different sources. Simply put, there could be several – or many – people in the world with identical handwriting, and there could be several – or many – riverbanks with identical mud, just as there may be more than one person in the world with the same DNA profile.

One only needs to mention these possibilities to indicate the difficulties that may be encountered in identifying suitable databases from which to generate reliable frequencies of occurrence for identical handwriting, chemically indistinguishable mud, etc. Setting those complications to one side, we can see that the version of illegitimately transposing the conditional known as source probability error can be, and perhaps frequently is, perpetrated in relation to a range of quantified and unquantified scientific and other expert evidence adduced in criminal proceedings.

The essential insight can be stated as a matter of logic without invoking any formal aspects of mathematics or probability calculations. A measure of similarity or "matching" simply cannot be equated with the likelihood of a common source.

3.25 **(e) *Underestimating the value of probabilistic evidence***

Illegitimately transposing the conditional typically makes the evidence in question appear stronger than it actually is. When it relates to prosecution evidence (as it frequently does), illegitimately transposing the conditional constitutes phoney proof of guilt, eroding and potentially undermining the presumption of innocence. There is, however, a complementary reasoning error which involves *under*valuing probabilistic evidence. This was dubbed "the defence attorney's fallacy" by Thompson and Schumann (1987), as a counterpoint to “the prosecutor’s fallacy”. Again, this terminology is not entirely apt and could mislead, because any participant in litigation, not only defence lawyers, might, in principle, undervalue evidence in this way. Moreover, “the defence attorney’s fallacy” is not a true mathematical fallacy (as the so-called prosecutor’s fallacy undoubtedly is), but rather a – conceptually speaking – straightforward misrepresentation of the value of probabilistic evidence.

3.26 Suppose that the frequency of blood type AB in a relevant population of 200,000 people is 1%. A suspect is found to have this blood type, matching blood recovered from a broken window at the scene of the crime. Intuitively, this is cogent – albeit not compelling – evidence linking the suspect to the crime-scene.

However, a sceptic might want to argue that the evidence has minimal probative value. The argument supposedly supporting this conclusion runs as follows. There are 200,000 potential suspects, and 2,000 of them would be expected (in the probabilistic sense) to have the blood type AB. If the suspect is merely one of 2,000 similarly situated individuals, the blood evidence might not be thought particularly probative against this, or any other, individual suspect. Indeed, it might now be argued that the evidence is insufficiently probative even to cross the minimal threshold of relevance to warrant legal admissibility. The evidence, it might be said, “proves nothing”.

3.27 Although “relevance”, “probative value”, and “proof beyond reasonable doubt” are indubitably different concepts that need to be carefully distinguished, both in theory and in practice, the sceptical conclusion is overstated. The figure 1/2,000 does *not* represent the value of the evidence of the matching blood type. It is perfectly true to say that, taken in isolation, the blood evidence (merely) places the suspect in a pool of 2,000 potential suspects. However, prior to obtaining the blood evidence the accused was in an undifferentiated pool of 200,000 suspects. The effect of the blood typing evidence is to narrow down that pool by a factor of 100, or in other words to increase the probability in favour of guilt by a factor of 100. Properly evaluated, the evidence is slightly over 100 times more likely if the suspect *is* the source of the blood on the broken window than if he is *not* the source (the probability of a match if the suspect is not the source is 1,999/200,000, or approximately 1/100). In summary, the figure of 100 is taken to represent the value of the evidence. This is powerful evidence, as we intuitively grasp.

Although it would not be capable of proving guilt beyond reasonable doubt if considered in isolation, its probative value is not fairly expressed by saying that the evidence "proves nothing". This interpretational error would be compounded if it were argued, more extravagantly still, that evidence of this kind should be excluded because it lacks sufficient probative value even to qualify as relevant evidence.

3.28 Proof of guilt is normally established, when it is, through a combination of different pieces of incriminating evidence. In Scotland, this expectation is formalised by a formal corroboration requirement necessitating independent evidence of the accused's guilt. Hence, the ultimate value of any particular piece of evidence, scientific or otherwise, must always be assessed contextually, in the light of its contribution to the case as a whole. This general precept is exemplified by the model jury direction suggested by the Court of Appeal in the well-known case of *Doheny and Adams*:

> "Members of the jury, if you accept the scientific evidence called by the Crown, this indicates that there are probably only four or five white males in the United Kingdom from whom that semen could have come. The defendant is one of them. If that is the position, the decision you have to reach, on all the evidence, is whether you are sure that it was the defendant who left that stain or whether it is possible that it was one of the other small group of men who share the same DNA characteristics".[21]

An unusual forensic application described in Gastwirth (1988), drawing on Usher and Stapleton (1979), arose in the following case.

> S, aged 16, became pregnant whilst a patient at a residential facility for those with severe mental disabilities. The pregnancy was terminated and the foetus examined to verify the most likely period of conception and to make serological tests. Because of the limited number (36) of men who possibly could have had access to S and the fact that about 90% of all men could be excluded based on appropriate tests, all 36 were asked to submit to serological tests and all agreed.

---

[21] *R* v *Doheny and Adams* [1997] 1 Cr App R 369, 375, CA. Also see *R* v *Lashley* (2000) and *R* v *Smith* (2000), discussed by Redmayne (2001: 74).

> The results of the test excluded all but four men and a further enzyme test excluded one more, reducing the potential list of suspects to three. These three included the police's prime suspect and another other two men regarded as "highly unlikely" to be the perpetrators. The prime suspect was another patient in the home whose disability was somewhat less severe than S's. The principal evidential value of the blood tests in this case was the elimination of innocent men from the list of suspects.

3.29 **(f) *Probability ("another match") error***

Two further quantities that are often confused in probabilistic reasoning are: (1) the frequency of an event within a designated population; and (2) the probability of a random match. This error might also potentially infect probabilistic evidence adduced in criminal proceedings, or its interpretation in criminal adjudication. It can be elucidated through a series of simple illustrations.

3.30 Suppose that a crime is committed, and evidence of a blood stain with a profile frequency of 1 in a million is found at the scene and identified as belonging to the offender. Consider the proposition that the evidence was *not* left by a particular suspect.

We know that the frequency of the profile of the stain is 1 in a million amongst the relevant population to which the offender is believed to belong. This means that if a person were chosen at random from that population the probability of that person's profile matching the profile of the blood stain is 1 in a million. This is the ***random match probability***. Notice, however, that this is *not* the same as saying that "the probability of finding another person in the population who has the same genetic profile is 1 in a million". In the first scenario, a person is chosen at random and a DNA profile obtained. The conclusion states the probability of achieving a match "in one go" (akin to the probability of choosing the ace of spades when making one draw from a shuffled standard deck of cards, i.e. 1/52). The second, "another match" probability relates to the occurrence of the event across an entire population, which for the ace of spaces in a standard deck is 1 (the card is definitely somewhere in the pack).

Consider a population comprising one million + 1 individuals, where the additional "+1" is the offender and there are one million innocent people. Then it can be calculated mathematically (see Appendix B) that the probability of at least one match with the offender amongst the one million innocent people is just over 3 out of 5 (0.63). This probability is obviously much larger than the profile frequency of 1 in a million.

3.31 The probability ("another match") error arises when the profile frequency of 1 in a million is equated to the probability of finding at least one other person in the population with the same frequency. A small value for the (random match) profile frequency is taken to imply a small value for the probability that at least one other person has the same matching profile. There is only one chance in 1 million that a person picked at random from the population shares a DNA profile that is common to one in every million people, but there is a 63% chance that there is at least one other person, somewhere, in a population comprising 1 million people who shares that profile.

This result is somewhat counter-intuitive, but it is plainly demonstrable. Consider a "population" of two fair coins, in which for each coin the probability of a head when the coin is tossed is p(head) = 0.5. The coins are secretly tossed once each; we do not know the outcome. Call a third coin, lying heads up, the *crime coin*. The issue is, will the population of tossed fair coins contain a match for the crime coin? The probability of observing at least one coin with a head ("another match") in the tossed coin "population" is not 0.5 (the random match probability for each coin), but 0.75. There are four, and only four, possible outcomes across the tossed coin "population": (a) the first coin is a head, the second coin is a head; (b) the first coin is a head, the second coin is a tail; (c) the first coin is a tail, the second coin is a head; (d) the first coin is a tail, the second coin is a tail. In three out of these four scenarios (75%, or 0.75) there is at least one head, matching the crime coin, in the tossed coin population. Only in scenario (d) is there no matching "head", giving a complementary probability of p(no match with crime coin) equal to ¼ = 25% = 0.25.

Analogously for the DNA profile example, probability ("another match") error is thinking that the probability of finding another person in the population of 1 million people (or 1 million secretly tossed coins) with the same genetic profile as the offender (crime coin) is 1 in a million. But the random match probability figure of 1 in a million is akin to the expected probability of tossing one coin and getting a head (0.5), as opposed to the probability of finding another person (tossed coin) in the population who has the same genetic profile (came up heads) as the offender (crime coin).

3.32 **(g) *Numerical conversion error***

Consider a characteristic which is prevalent in only 1 in a thousand, 1/1,000, people (e.g. a height greater than a certain designated value, such as two metres). It is sometimes claimed that the significance of evidence of this characteristic can be expressed in terms of the number of people who would have to be counted before there is another (random) match, being the reciprocal of the frequency (1,000, in this example); i.e. 1,000 people would need to be observed before someone else of that height would be encountered. This is an obvious fallacy, since the very next person observed could be that height or taller.

A frequency of 1/1,000 does not mean that a match (with heights, as in this example, or with any other designated characteristic) is expected only on every thousandth experimental observation. This would almost be like saying that, if one in every thousand motorists will cause a serious accident, we should confiscate the licences of every thousandth driver we encounter. Numerical conversion error featured in the American case of *Ross* v *State*.[22] The relative frequency of a DNA profile was calculated as 1 in 23 million. On the strength of this calculation, the expert testified that he would not expect to encounter another individual with that profile until testing at least another 23 million people. This considerably exaggerates the probative value of a matching DNA profile. It can be calculated mathematically that, for a relative frequency of 1 in 23 million, just under 16 million people would need to be tested in order to achieve a probability of at least 0.5 ("as likely as not") of identifying someone other than the defendant with that profile.

3.33 **(h) *False positive probability (distinguished from the probability that a declared match is false)***

[22] *Ross* v *State*, Court of Appeals of Texas, Houston (14th Dist.) 13 February 1992, transcript quoted by Koehler (1993: 34).

Serious errors of interpretation can occur through ignorance or underestimation of the potential for a ***false positive***. A false positive result in a scientific or medical test, for example, is one in which the test gives a positive result indicating the presence of the substance or disease for which the test was conducted when, in reality, that substance or disease is not present. In contrast, a ***false negative*** result is one in which the test gives a negative result indicating the absence of the substance or disease, etc. for which the test was conducted when in fact the substance or disease is present.

Many types of scientific and other expert evidence adduced in criminal proceedings have the potential for generating false positives (and false negatives). For example, a forensic scientist might declare "a match" between a DNA profile taken from a crime scene and a DNA profile from a suspect. Suppose, in reality, the suspect does not have the same profile as the perpetrator nor is he the source of the crime scene stain. The result is a false positive. Reported matches relating to fingerprints, ballistics, and various forms of trace evidence (blood, semen, hairs, fibres, firearms residue, etc.), amongst others, are likewise susceptible to false positives (reported matches, where there is no match in fact). The ***false positive probability*** is the probability of reporting a match when the suspect and the real perpetrator do not share the same DNA profile, or where the suspect's and crime-scene fingerprints, blood, fibres or whatever do not, in fact, match.

3.34 Once again, it is vital to pay close attention to the precise wording of these expressions (that is, to specify the precise question which the evidence is being adduced to answer) and to be on one's guard against illegitimate conflations of quite different quantities. Here, in particular, it would be fallacious to equate a value for the ***false positive probability*** (the prior probability of declaring a match falsely) with the value for the ***probability of a false match*** (the probability that any given *declared* match is false). Despite the linguistic similarity of these formulations, they represent categorically different concepts of probability. The first value is a measure of the reliability of testing procedures, which is given by the percentage of non-matches reported as matches (the frequency of reported matches that are not true matches); the second value is the probability that, a match having been declared, it will be a false match. The probability of a false positive is the probability of a match being reported under a specified condition (no match). It does not depend on the probability of that condition occurring, since the condition (no match) is already assumed to have occurred. By contrast, the probability that the samples do not match when a match has been reported depends on *both* the probability of a match being reported under the specified condition (no match) *and* on the prior probability that that condition will occur. Consequently, the probability that a reported match is a true match or a false match cannot be determined from the false positive probability alone.

The distinction between false positive probability and the probability that a declared match is false has important implications for interpreting the reliability and probative value of scientific evidence. A particular laboratory may have a low false positive rate in the sense that it does not often report false matches. However, this does not necessarily mean that when the laboratory declares a match there is a high probability that it is a true match rather than a false positive. The probability that a declared match is a false positive is partly determined by pertinent base rates, which can have unanticipated effects (as we saw in the Blue and Red Buses hypothetical discussed in §2.30–§2.31). The following pair of hypothetical illustrations should serve to reinforce the message.

3.35 Suppose that, in a relevant population of 10,000 individuals, the base-rate for Disease X is 1% (100 people). A person chosen at random from the population therefore has a probability of 0.01 of being infected. The probability that a particular diagnostic test for the disease will give a positive result if a person has the disease is known to be 0.99. So for the 100 people that actually have the disease, 99 will give a positive test result. A negative result would be recorded for the other infected individual, who is the one ***false negative***.

The probability that this same diagnostic test will give a negative result if a person does not have the disease is stipulated to be 0.95. Thus, for the 9,900 people who do not have the disease, 9,405 would give a negative test result. The other 495 people will test positive, even though they do not actually have the disease. They are ***false positives*** and the ***false positive probability*** is 0.05 (5%). Employing the terminology of "sensitivity" and "specificity" introduced in §2.21, we can say that the sensitivity of the diagnostic test is 0.99, and its specificity is 0.95.

These results are summarised in the following table:

Table 3.1: Results of a Diagnostic Test for Disease X

	Diagnostic Test		Total
	Positive	Negative	
Disease X present	99	1	100
Disease X absent	495	9,405	9,900
Total	594	9,406	10,000

Suppose that an individual tests positive for Disease X. What is the probability that this person actually has the disease?

From the table, we can clearly see that the number of people expected to test positive for the disease is 594. Of those 594 people, 99 will actually have the disease. Thus, the probability that a person with a positive result for the test actually has the disease is 99/594 = 1/6. Complementarily, the probability that a person with a positive test result does not have the disease is 495/594 = 5/6.

The diagnostic test is both highly sensitive and highly specific to Disease X, generating an intuitive expectation that the test should be highly reliable. However, because the base rate for the disease in the population is very low (1%) the probability of a declared match being false is surprisingly high – 495/594 = 5/6. The probability that a declared match is a false positive is completely different to the ***false positive probability*** for the diagnostic test, which is a measure of the test's specificity. From the table, we can see that the test will incorrectly diagnose 495 out of the 9,900 people in the population who are not infected with Disease X, i.e. 495/9,000 = 0.05; which is the complement of the test's stipulated specificity (0.95). The probability that a declared match is false varies with changes in the base rate (and at the limit, if the base rate were zero the probability that a declared match is false would be 1, and vice versa), whereas the specificity of a diagnostic test is unaffected by changes in the base rates for infection.

3.36 A second hypothetical example using the same numbers but this time referring to DNA evidence will clarify the significance of this distinction for criminal proceedings.

Table 3.2: Results of DNA Profiling

	DNA Evidence		Total
	Present	Absent	
Guilty	99	1	100
Innocent	495	9,405	9,900
Total	594	9,406	10,000

Consider Table 3.2. In this variation, the prior probability of guilt (base rate) is 1% (100/10,000); the probability that the evidence is detected on a person who is guilty is 0.99 (99/100); the probability the evidence is absent on a person who is innocent is 0.95 (9,405/9,900). The number of people on whom the evidence is present is 594, of whom 99 are guilty. The other 495 on whom the evidence is detected are innocent false positives. Thus, the probability that person on whom the evidence is detected is guilty is 99/594 = 1/6.

The *false positive fallacy* (Thomson *et al* 2003) is to equate the antecedent probability of a false positive (presence of the evidence when a person is innocent) with the probability that a person on whom the evidence is present is nonetheless innocent. In this illustration:

> (i) the probability of a false positive is 495/9,900 = 1/20 = 0.05 (in other words, the test is 95% specific for matching DNA profiles);
>
> (ii) the probability a person is innocent when the evidence is present (a match has been declared for the DNA profiles) = 495/594 = 5/6 = 0.833 (approx.).

The second probability is obviously much larger (and the corresponding event more likely) than the first, and it would be a serious error to confuse them with each other.

3.37 **(i) *Fallacious inferences of certainty***

A very low probability of a random match is sometimes thought to equate to a unique identification. For example, a DNA profile with a very small random match probability might be taken to imply that the possibility of encountering another person living on earth with the same DNA profile is effectively zero; in other words, that there is sufficient uniqueness within the observed characteristics to eliminate all other possible donors in the world. Influenced by similar thinking, the US Federal Bureau of Investigation decided that FBI experts could testify that DNA from blood, semen, or other biological crime-stain samples originated from a specific person whenever the random match probability was smaller than 1 in 260 billion (Holden, 1997).

3.38 However, all such inferences of uniqueness are epistemologically unwarranted. Probabilistic modelling must be adjusted to accommodate the empirical realities of criminal proceedings. For example, there may be contrary evidence, such as an alibi, or risks of contamination of samples, etc. Also, some of the modelling assumptions underpinning the probabilistic calculations may be open to challenge. In the final analysis, no probability of any empirical event (e.g. the probability of another person matching a DNA profile), however small, can be equated to a probability of zero (no person with a matching profile living anywhere in the world). Even though a random match probability may be extremely small (one in ten billion, say – the world's estimated current population being (only) six billion) it does not warrant the inference that a matching DNA profile uniquely identifies an individual. Quite apart from anything else, every set of identical twins in the world has the same DNA profile – and the chances of obtaining random matches are vastly increased in relation to parents and siblings.

With a random match probability of, e.g., one in ten billion and a world population of six billion, the probability that there is at least one other person with the profile is about 0.46 (and a corresponding probability of 0.54 that no-one else does). For six billion people and a random match probability of 1 in 260 billion, the probability of at least one other match in the population is about 0.02.

3.39 There appears to be growing sophistication in probabilistic reasoning across the forensic sciences, which has been spearheaded by developments in DNA profiling. Commenting on this trend, Saks and Koehler (2005) anticipate "a paradigm shift in the traditional forensic identification sciences" suggesting that "the time is ripe for the traditional forensic sciences to replace antiquated assumptions of uniqueness and perfection with a more defensible empirical and probabilistic foundation". The idea here is that DNA evidence and the probabilistic techniques applied to it will become a kind of "gold standard" for all forensic science evidence. DNA evidence will be explored at greater length in *Practitioner Guide No 2*.

3.40 **(j) *Unwarranted assumptions of independence***

Probabilistic concepts of independence and dependence were introduced in Section 2 of this *Guide*. Our final "trap for the unwary" involves assuming that two probabilities are independent, and therefore amenable to the ***product rule for independent events***, when that assumption is unwarranted. Either known information demonstrates that the two events are related, or there are insufficient data to make any reliable assumption either way (and therefore the default assumption should be dependence in criminal proceedings).

3.41 A real-world illustration of fallacious assumptions of independence is afforded by *Sally Clark*'s case.[23] Research data showed that the frequency (probability) of sudden infant death syndrome (SIDS) in a family like the Clarks' was approximately 1 in 8,543. From this it was deduced, applying the product rule for independent events, that the probability of two SIDS deaths in the same family would be 1/8,543 x 1/8,543 = 1/72,982,849, which was rounded down to produce the now notorious statistic of "1 in 73 million" quoted in court. The fact-finder was apparently encouraged to believe that the figure of 1 in 73 million implied that multiple SIDS deaths in the same family would be expected to occur about once every hundred years in England and Wales. Of course, this calculation and deduction are valid only on the assumption that two SIDS deaths in the same family are entirely unrelated, independent, events. But this was a perilously fallacious assumption.

In reality, the assumption of independence was directly contradicted by the research study from which the original 1/8,543 statistic was derived. Fleming *et al* (2000) reported that a sibling had previously died and the death ascribed to SIDS in more researched SIDS families than in control sample families (1.5%, five out of 323 families, and 0.15%, two out of 1288 families, respectively, and that these percentages were significantly different in the statistical sense). Far from warranting an assumption of independence, these empirical data suggest that multiple SIDS in the same family may be dependent events.

3.42 Recall that interpretation of evidence is a fundamentally comparative exercise. The true probative value of evidence can be assessed only by considering it under at least two propositions, which in criminal proceedings can be modelled as "the proposition advanced

[23] *R* v *Clark* [2003] EWCA Crim 1020.

by the prosecution" and the competing "proposition advanced by the defence" (which, in the absence of anything more suitable, may simply be the negation of the prosecution's proposition).

When the evidence is implausible under the defence proposition, it is tempting to jump to the conclusion that the prosecution's case (proposition) must be true. But that inference is speciously premature. The evidence might be *even more* implausible assuming the truth of the prosecution's proposition. For example, it might be very unlikely that two cases of SIDS would be experienced in a single family. But it might be even less likely that a mother would serially murder her two children (we must make assumptions here, of course, about the impact of other evidence). So, taken in isolation, the bare fact of two infant deaths in the same family is probably more likely to be SIDS than murder. Unlikely though the former innocent explanation may be, it is not as unlikely as the latter, incriminating explanation.

3.43 Forensic scientists and other expert witnesses in criminal proceedings should guard against making unwarranted assumptions of independence. That two events or characteristics are truly independent should be demonstrated rather than merely assumed before applying the product rule for independent events to calculate the probability of their conjunction. Witnesses who testify on the basis of independence should be prepared to explain and justify their rationale for that supposition, whilst lawyers should be ready to probe statements of the form "research shows that…" in order to satisfy themselves that the quoted research is fit for purpose and that the evidence does not rest on unwarranted assumptions of independence.

Appendix B – Technical Elucidation and Illustrations

Sample Size and Percentages

Sample size is important when considering the precision of estimates. Consider an experimental trial like the example given in §2.7. The sample comprised 1,000 spins of a standard roulette wheel. In percentage terms, the difference between the expected and observed frequencies of the ball landing in the no.1 slot was calculated to be 0.8%; the difference in the absolute frequencies was 35 (observed) to 27 (expected) no.1 slots. Trials comprising 10,000 spins or only 100 spins, however, would be expected to produce, respectively, more or less reliable estimates. As a rule of thumb, the precision of an estimate is related to the square root of the sample size; in order to double the precision of an estimate it is necessary to quadruple the sample size.

Consider another illustration based on coin-tossing. Thirteen heads in twenty tosses of a fair coin (65% heads) is not unusual; using standard probabilistic calculations thirteen or more heads would be expected to occur once in every seven or eight sets of 20 tosses of a fair coin. However, 130 heads in 200 tosses of a fair coin (also 65% heads) *would* be unusual – 130 or more heads would be expected about once in every 550 sets of 200 tosses of a fair coin..

The Multiplication (Product) Rule for Probability[24]

The multiplication rule for probability concerns the conjunction of events. It is best introduced through an artificial example. Consider an urn containing black and white balls in proportions b and w, respectively, where proportions are taken to be numbers between 0 and 1, and b and w are such that $b + w = 1$. The exact number of balls of each colour is not important. In addition to the colour of the balls, assume each ball is either spotted or plain with proportions s and p, and where $s + p = 1$. There are then four types of ball: 'black, spotted', 'black, plain', 'white, spotted' and 'white, plain', denoted c, e, d and f, respectively, such that $c + d + e + f = 1$; $c + d = s$; $e + f = p$; $c + e = b$; and $d + f = w$. These results are conveniently displayed in Table B1.

[24] This section draws on Lindley (1991).

Table B1: Proportions of black, white, spotted and plain balls in an urn

	Black	White	Total
Spotted	c	d	s
Plain	e	f	p
Total	b	w	1

The proportions of spotted and plain balls (*s* and *p*) are given in the final column, labelled 'Total'. The proportions of the black and white balls (*b* and *w*) are given in the final row, also labelled 'Total'.

Let *K* denote the composition of the urn. Let *B* be the event that a ball drawn at random is black and *S* be the event that a ball drawn at random is spotted. Thus, the event that a ball drawn at random is black and spotted is denoted '*B* and *S*'. For conjunctions, the 'and' is often dropped. In this example '*B* and *S*' would be written as *BS*. Proportions can easily be translated into probabilities, since they obey the same rules of logic. Thus, the probability that a ball drawn at random is black, given the composition *K* of the urn, is *b*. Similarly, the probability a ball drawn at random is spotted, given the composition of the urn, is *s*. The probability a ball drawn at random is spotted and black is *c*.

A new idea is now introduced. Suppose someone else had withdrawn a ball at random and announced, truthfully, that it was black. What is the probability that this black ball is also spotted? It is equivalent to the proportion of spotted balls which are also black, which from Table B1 is *c/b*, spotted over black.

Consider the trivial result that

$$c = b \times (c/b).$$

In words, the proportion *c* of balls that are both black and spotted is the proportion *b*, balls that are black, multiplied by the proportion of spotted balls amongst the black balls (*c* out of *b*, or *c/b*).

The equivalent result for probabilities is

$$p(B \text{ and } S) = p(B) \times p(S \mid B).$$

Section 2.35 gives an example of this result applied to the drawing of Aces without replacement from a pack of playing cards. Event B is the drawing of an Ace in the first draw, event S is the drawing of an Ace in the second draw. The left-hand-side of the equation is the drawing of two Aces, which was shown by direct enumeration to have a probability of 1/221. For the right-hand-side, $p(B) = 1/13$ and $p(S \mid B)$ is the probability of drawing an Ace as the second card given that an Ace has been drawn as the first card, which has been shown to be 1/17. The product of 1/13 and 1/17 is 1/221, which is equal to the value on the left-hand-side.

Conditional Probabilities for Dependent Events – A Counter-intuitive Result

One might anticipate that the conditional probability of two dependent events would always be smaller than the probability of the first event taken in isolation. For example, the probability of drawing an Ace from a normal playing deck is 4/52 = 1/13, whereas the probability of drawing an Ace after an Ace has already been drawn without replacement is 3/51 = 1/17. The probability of drawing an Ace after *two* Aces have already been drawn without replacement is even smaller, 2/50 = 1/25.

However, in some cases the probability of an event conditional on another event is actually *greater than* the unconditional probability of the event. Imagine that the frequency of baldness in the general population is 10%. The probability that a person selected at random is bald is therefore 0.10. But notice how these probabilities change if we condition the probability of baldness on gender. Now we would intuitively expect the frequency of baldness *conditioned on being male* to increase, say to 20%; and the frequency of baldness *conditioned on being female* to decrease, say to (almost) 0%. Conditioned on gender, the probability that a person selected at random who is male is also bald is 0.20. And the probability that a person selected at random who is female is also bald is nearly zero. So the frequency of baldness conditioned on gender may be *greater or less* than the unconditional population frequency of baldness.

This result is obtained only for *dependent events*, as where maleness also predicts baldness. If one were to assume independence of baldness and gender, the probability that a person selected at random from the population is bald would remain 0.10 as before, regardless of whether that probability were conditioned on the person's being male, or female, or of unknown gender.

For dependent events only, a conditioning event (gender in the example) may cause the probability of the original event (baldness) to increase or decrease, depending on the nature of the conditioning event.

Interrogating Base Rates

Statistical data, such as those adduced in criminal proceedings as base rates (see §§2.20-2.22, above), need to be interpreted with care. A statistic expressed as a percentage or ***relative frequency*** may be entirely valid, in a formal sense, and yet still potentially seriously misleading. Kaye and Freedman (2000), in their contribution to the US Federal Judicial Center's *Reference Manual on Scientific Evidence*, identify a number of pertinent questions that one might ask when interrogating base rates:

1. *Have appropriate benchmarks been provided?*

   Selective presentation of numerical information can be misleading. Kaye and Freedman (2000) cite a television commercial for a mutual fund trade association which boasted that a $10,000 investment in a mutual trade fund made in 1950 would have been worth $113,500 by the end of 1972. However, according to the *Wall Street Journal*, that same $10,000 investment would have grown to $151,427 if it had been spread over all the stocks comprising the New York Stock Exchange Composite Index.

2. *Have data collection procedures changed?*

   One of the more obvious pitfalls in comparing data time series is that the protocols for data collection may have changed over time. For example, apparent sharp rises or falls in social data, such as morbidity or crime rates, may be mere artefacts of changes in data reporting or recording practices with absolutely no bearing on the underlying social reality.

3. *Are data classifications appropriate?*

   Data can be classified and organised in different ways. One must therefore be alive to the possibility that a particular classification has been selected quite deliberately to support a particular argument or to a highlight a favourable comparison – and by implication to downplay unfavourable arguments or comparisons. Gastwirth (1988b) cites the following example from the USA.

In 1980, tobacco company M sought an injunction to stop the makers of T low-tar tobacco from running advertisements claiming that participants in a national taste test preferred T to other brands. The plaintiffs objected that the advertising claims that T was a "national test winner" and "beats" other brands were false and misleading. In reply, the defendant invoked the data summarised in Table B2 as evidence.

Table B2: The preferences of participants in a national taste test for the comparison of T and M tobacco.

	T much better than M	T somewhat better than M	T about the same as M	T somewhat worse than M	T much worse than M
Number	45	73	77	93	36
Percentage	14	22	24	29	11

According to these data, more survey respondents judged T much better than M (14%) than those finding T much worse than M (11%). Also, 60% regarded T as better or the same as M (i.e. including the 24% who expressed no preference either way). But another way of interpreting these data is to note that 40% who expressed a clear preference actually preferred M to T, whilst only 36% actively preferred T to M. The court ruled in favour of the plaintiffs.

4. *How big is the base of a percentage?*

When the base is small, actual numbers may be more informative than percentages. For example, an increase form 10 to 20 and an increase from 1 million to 2 million are both 100% increases. To say that something has increased "by 100 per cent" always sounds impressive, but whether it is or not depends, amongst other things, on the numbers behind the percentage. (Also recall the coin-tossing examples of 13 heads in 20 tosses and 130 heads in 200 tosses, discussed in the first section of this Appendix.)

5. *Which comparisons are made?*

Comparisons are always made relative to some base-line, so that the choice of base-line (where eligible alternatives are available) may be a crucial factor in interpreting the meaning of any statistic. Suppose that a University reports that the proportion of first class degrees awarded in humanities subjects has increased by 30% on the previous year. All well and good. But is the previous year an appropriate base-line? What if the previous year was a markedly fallow year for first class degrees in the humanities, so that a 30% increase merely restores the level of firsts to what it was two years ago? Conversely, there may have been a big increase in firsts in the previous year as well, perhaps suggesting a worrying erosion in academic standards rather than an impressive improvement in student performance. In this and many other similar scenarios, choice of base-line has a major bearing on the meaning – and probative value – of statistical information.

Illegitimately transposing the conditional – case illustrations

There are numerous reported cases involving illegitimate transpositions of the conditional ("***the prosecutor's fallacy***"). This is how it occurred in *Deen*[25] in relation to a DNA profile with a frequency of 1 in 3 million in the relevant population:

> *Prosecuting counsel*: So the likelihood of this being any other man but Andrew Deen is one in 3 million?
>
> *Expert*: In 3 million, yes.
>
> *Prosecuting counsel*: You are a scientist... doing this research. At the end of this appeal a jury are going to be asked whether they are sure that it is Andrew Deen who committed this particular rape in relation to Miss W. On the figure which you have established according to your research, the possibility of it being anybody else being one in 3 million what is your conclusion?
>
> *Expert*: My conclusion is that the semen originated from Andrew Deen.
>
> *Prosecuting counsel*: Are you sure of that?
>
> *Expert*: Yes.

---

[25] *R* v *Deen*, CA, *The Times*, 10 January 1994.

The fallacy is perpetrated when the expert is induced to agree that the likelihood (probability) of the criminal being someone other than Andrew Deen, given the evidence of the DNA match, is one in three million. (This error was further compounded by the unwarranted source-level conclusion that Deen was the source of the stain, i.e. ***source probability error***.)

The relative frequency of the DNA profile in the relevant population was 1 in 3 million, meaning that one person in every 3 million selected at random from this population would be expected to have a matching profile. This is patently not the probability that a person with a matching profile is innocent, as the quoted exchange between the expert and prosecuting counsel clearly implies. The conditional has been transposed illegitimately. One cannot calculate the probability of guilt or innocence of a particular person without knowing the number of people in the relevant suspect population. If the suspect population comprised, say, 6 million individuals, one would expect two matching profiles amongst the innocent people. Add this to the offender (whose probability of matching can be taken to be 1) and the expected number of people with the profile is 3, giving a probability of guilt for a person with the profile – $p(G|E) = 1/3$.

An expert witness called by the prosecution also illegitimately transposed the conditional in *Doheny and Adams*, as recounted by the Court of Appeal:[26]

> "A. Taking them all into account, I calculated the chance of finding all of those bands and the conventional blood groups to be about 1 in 40 million.
> Q. The likelihood of it being anybody other than Alan Doheny?
> A. Is about 1 in 40 million.
> Q. You deal habitually with these things, the jury have to say, of course, on the evidence, whether they are satisfied beyond doubt that it is he. You have done the analysis, are you sure that it is he?
> A. Yes."
> The question, in leading form, and the numerical answer given to it constituted a classic example of the 'prosecutor's fallacy'. The third question was one for the jury, not for the witness. The witness gave an affirmative answer to it. It is not clear to what evidence, if any, other than the DNA evidence, he had regard when giving that answer. For the reasons that we gave in our introduction to this Judgment, this series of questions and answers was inappropriate and potentially misleading.

[26] *R* v *Doheny and Adams* [1997] 1 Cr App R 369, 377-8, CA.

A third illustration comes from *Gordon*,[27] where the relative frequencies of the DNA profiles in question were calculated to be 1 in ten-and-a-half million and 1 in just over seventeen million. An expert witness testified that 'she was sure of the match between the semen samples and the appellant's blood'.[28] This is ***source probability error***, since even the extreme unlikelihood of a random match does not permit the expert to infer a definitive source. Fundamentally, to confuse the probability that a DNA profile derived from a crime scene will match an innocent person's profile (the ***random match probability***) with the probability that a person with a matching profile is innocent, as the expert appears to have done in *Gordon*, is to commit the fallacy of illegitimately transposing the conditional.

Calculating the probability of "another match"

As we explained in §, the probability of finding "another match" should not to be confused with the ***random match probability***. Here is the more technical explanation.

Consider a characteristic which is prevalent in only 1 in a thousand, 1/1,000, people (e.g. a height greater than a certain designated value, such as two metres). It is sometimes claimed that the significance of evidence of this characteristic can be expressed in terms of the number of people who would have to be counted before there is another (random) match, being the reciprocal of the frequency (1,000, in this example); i.e. "1,000 people would need to be observed before someone else of that height would be encountered". Yet this is an intuitively obvious fallacy, since the very next person observed could be that height or taller.

This result can be demonstrated formulaically. It has been established that the probability that a person is no taller than two metres is 999/1,000. If n independent (unrelated) people are observed, we also know by repeated use of ***the product rule for independent events*** that the probability that none is taller than two metres is $(999/1000)^n$ (the probability is 999/1000 on each selection, and we make n independent selections). The complementary event is that at least one person is taller than two metres in height, i.e. $1 - (999/1000)^n$. For it to be more likely than not that at least one person is taller than two meters, 1 -

[27] *R* v *Gordon* [1995] 1 Cr App R 290, CA.

[28] ibid. 293.

$(999/1000)^n$ must be greater than 0.5. In fact the formula $1 - (999/1000)^n$ equals 0.5 when $n = 693$, so it is more likely than not that at least one person will be taller than two metres after selecting 694 people – *not* after 1,000 selections. If 1,000 people were indeed observed, the probability that at least one of them would be over two metres in height is 0.632. In order to raise the probability of at least one other person of at least that height to 0.9 one would need to look at 2,307 people, which is the value of n where $1 - (999/1000)^n = 0.9$.

General Principles for the Presentation of Scientific Evidence

Various attempts have been made over the years to formulate general principles to guide the presentation and interpretation of scientific and other expert evidence in criminal proceedings. Here, for ease of reference, we summarise two significant sources of normative guidance.

First, Part 33 (Expert Evidence) of the Criminal Procedure Rules 2010 includes the following requirements:

Rule 33.2 - Expert's duty to the court

(1) An expert must help the court… by giving objective, unbiased opinion on matters within his expertise.

(2) This duty overrides any obligation to the person from whom he receives instructions or by whom he is paid.

(3) This duty includes an obligation to inform all parties and the court if the expert's opinion changes from that contained in a report served as evidence or given in a statement.

Rule 33.3 - Content of expert's report

(1) An expert's report must—

(a) give details of the expert's qualifications, relevant experience and accreditation;

(b) give details of any literature or other information which the expert has relied on in making the report;

(c) contain a statement setting out the substance of all facts given to the expert which are material to the opinions expressed in the report, or upon which those opinions are based;

(d) make clear which of the facts stated in the report are within the expert's own knowledge;

(e) say who carried out any examination, measurement, test or experiment which the expert has used for the report and—

(i) give the qualifications, relevant experience and accreditation of that person,

(ii) say whether or not the examination, measurement, test or experiment was carried out under the expert's supervision, and

(iii) summarise the findings on which the expert relies;

(f) where there is a range of opinion on the matters dealt with in the report—

(i) summarise the range of opinion, and

(ii) give reasons for his own opinion;

(g) if the expert is not able to give his opinion without qualification, state the qualification;

(h) contain a summary of the conclusions reached;

(i) contain a statement that the expert understands his duty to the court, and has complied and will continue to comply with that duty; and

(j) contain the same declaration of truth as a witness statement.

These criteria for expert report writing may be regarded *mutatis mutandis* as general expectations of scientific evidence adduced in legal proceedings in any form, including live oral testimony. The Court of Appeal has reiterated the vital importance of full compliance with CrimPR 2010 Rule 33 on many occasions.

Further normative guidance might be found in the following list of criteria and associated principles, which have been advanced by the Association of Forensic Science Providers:[29]

- *Balance*: The expert should address at least one pair of propositions.
- *Logic*: The expert will address the probability of the evidence given the proposition and relevant background information and not the probability of the proposition given the evidence and background information.
- *Robustness*: The expert will provide factual and opinion evidence that is capable of scrutiny by other experts and cross-examination. Expert evidence will be based on sound knowledge of the evidence type(s) and use verified databases, wherever possible.

[29] The Association of Forensic Science Providers aims to "represent the common interests of the providers of independent forensic science within the UK and Ireland with regard to the maintenance and development of quality and best practice in forensic science and expert witness in support of the Justice System, from scene to court, irrespective of the commercial pressures associated with the competitive forensic marketplace": see Brown and Willis (2009); Association of Forensic Science Providers (2009).

- *Transparency*: The expert will be able to demonstrate how inferential conclusions were produced: propositions addressed, examination results, background information, data used and their provenance.

These desiderata for expert evidence encapsulate several of the points stressed in this Report. The first principle expresses the idea that it is not sufficient to consider the value of evidence – even strongly incriminating evidence – in the abstract. Evidential value is a function of two competing propositions, the likelihood of the evidence on the assumption that the prosecution's proposition is true and the likelihood of the evidence on the assumption that the prosecution's proposition is false. The second principle reiterates the elementary injunction against illegitimately transposing the conditional. As a general rule, forensic scientists and other expert witnesses should be assessing the probability of the evidence, rather than commenting on the probability of contested facts (much less the ultimate issue of guilt or innocence). Robustness is concerned with scientific methodology, which must be valid and able to withstand appropriately searching scrutiny. The knowledge of the expert must be sound. Laboratory equipment must be in good working order, properly calibrated. Operational protocols should be validated with known error rates. Databases will have been verified or accredited as much as possible. Finally, the principle of transparency states that all of the assumptions, data, instrumentation and methods relied on in producing the evidence must stated explicitly or at least open to examination and verification by the court.

Appendix D – Select Bibliography

Aitken, C.G.G. and Taroni, F. (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester: Wiley.

- (2008) 'Fundamentals of Statistical Evidence – A Primer for Legal Professionals' 12 *International Journal of Evidence & Proof* 181.

Allen, R.J. (1991) 'The Nature of Juridical Proof' (1991) 13 *Cardozo LR* 373.

Allen, R.J. and Pardo, M. (2007) 'The Problematic Value of Mathematical Models of Evidence' 36 *Journal of Legal Studies* 107.

Allen, R.J. and Redmayne, M. (eds.) (1997) *Special Issue on Bayesianism and Juridical Proof* 1(6) *International Journal of Evidence & Proof* 253.

Allen, R.J. and Roberts, P. (eds.) (2007), *Special Issue on the Reference Class Problem* 11(4) *International Journal of Evidence & Proof* 243.

Association of Forensic Science Providers (2009) 'Standards for the Formulation of Evaluative Forensic Science Expert Opinion' 49 *Science and Justice* 161.

Balding, D.J. (2005) *Weight-of-Evidence for Forensic DNA Profiles*. Chichester: Wiley.

Balding, D.J. and Donnelly, P. (1994) 'The Prosecutor's Fallacy and DNA Evidence' *Criminal Law Review* 711.

Brown, S. and Willis, S. (2009) 'Complexity in Forensic Science' 1 *Forensic Science Policy and Management* 192.

Buckleton J.S. (2004) 'Population Genetic Models' in J.S. Buckleton, C.M. Triggs and S.J. Walsh (eds.) *DNA Evidence*. Boca Raton, Florida: CRC Press.

Callen, C.R. (1982) 'Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law' 57 *Indiana Law Journal* 1.

- (1991) 'Adjudication and the Appearance of Statistical Evidence' 65 *Tulane Law Review* 457.

Champod C., Evett I.W. and Jackson, G. (2004) 'Establishing the Most Appropriate Databases for Addressing Source Level Propositions' 44 *Science and Justice* 153.

Coleman, R.F. and Walls, H.J. (1974) 'The Evaluation of Scientific Evidence' *Criminal Law Review* 276.

Cook, R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. (1998a) 'A Model for Case Assessment and Interpretation' 38 *Science & Justice* 151.

- (1998b) ‘A Hierarchy of Propositions: Deciding which Level to Address in Casework’ 38 *Science & Justice* 231.

- (1999) ‘Case Pre-assessment and Review of a Two-way Transfer Case’ 39 *Science & Justice* 103.

Dawid, A. P. (2005) ‘Probability and Proof’, on-line Appendix I to T.J. Anderson, D.A. Schum and W.L. Twining, *Analysis of Evidence: Second Edition*. Cambridge: CUP. http://tinyurl.com/7g3bd (accessed 19 October 2010).

DeGroot, M. H., Fienberg, S.E. and Kadane, J.B. (eds.) (1994) *Statistics and the Law*. New York: Wiley.

Diamond, S.S. (2000) ‘Reference Guide on Survey Research’ in *Reference Manual on Scientific Evidence*, 2nd edn. Federal Judicial Center: Washington, DC.

Eggleston, R. (1983) *Evidence Proof and Probability,* 2nd edn. London: Butterworths.

Evett, I.W., Foreman, L.A., Jackson, G. and Lambert, J.A. (2000) ‘DNA Profiling: A Discussion of Issues Relating to the Reporting of Very Small Match Probabilities’ *Criminal Law Review* 341.

Evett, I.W., Jackson, G., Lambert, J.A. and McCrossan, S. (2000) ‘The Impact of the Principles of Evidence Interpretation and the Structure and Content of Statements’ 40 *Science & Justice* 233.

Evett, I.W. and Weir, B.S. (1998) *Interpreting DNA Evidence.* Sunderland, Mass.: Sinauer Associates Inc.

Fienberg, S. E. (ed.) (1989) *The Evolving Role of Statistical Assessments as Evidence in the Courts*. New York: Springer.

Finkelstein, M. (2009) *Basic Concepts of Probability and Statistics in the Law*. New York: Springer.

Finkelstein, M.O. and Levin, B. (2001) *Statistics for Lawyers*, 2nd edn. New York: Springer.

Fleming, P., Blair, P., Bacon, C., Berry, J. (2000) *Sudden Unexpected Deaths in Infancy.* London: HMSO.

Friedman, R.D. (1996) ‘Assessing Evidence’ 94 *Michigan Law Review* 1810.

Gastwirth, J.L. (1988a) *Statistical Reasoning in Law and Public Policy, vol 1: Statistical Concepts and Issues of Fairness*. Boston, Mass.: Academic Press.

- (1988b) *Statistical Reasoning in Law and Public Policy, vol 2: Tort Law, Evidence and Health*. Boston, Mass.: Academic Press.

- (ed.) (2000) *Statistical Science in the Courtroom*. New York: Springer.

Hodgson, D. (1995) ‘Probability: The Logic of the Law – A Response’ 15 *Oxford Journal of Legal Studies* 51.

Holden, C. (1997) ‘DNA Fingerprinting Comes of Age’ 278 *Science* 1407.

Jackson, G. (2009) ‘Understanding Forensic Science Opinions’ in J. Fraser and R. Williams (eds.), *Handbook of Forensic Science*. Cullompton, Devon: Willan Publishing.

Jackson, G., Jones, S., Booth, G., Champod, C. and Evett, I.W. (2006) ‘The Nature of Forensic Science Opinion - A Possible Framework to Guide Thinking and Practice in Investigations and in Court Proceedings’ 46 *Science and Justice* 33.

Kadane, J.B. (2008) *Statistics in the Law: A Practitioner’s Guide, Cases, and Materials*. New York: OUP.

Kaye, D.H. (1979) ‘The Laws of Probability and the Law of the Land’ 47 *University of Chicago Law Review* 34.

- (1984) ‘Thinking Like a Statistician: The Report of the American Statistical Association Committee on Training in Statistics in Selected Professions’ 34 *Journal of Legal Education* 97.

- (1993) ‘DNA Evidence: Probability, Population Genetics and the Courts’ 7 *Harvard Journal of Law and Technology* 101.

Kaye, D.H. and Freedman, D.A. (2000) ‘Reference Guide on Statistics’ in *Reference Manual on Scientific Evidence*, 2nd edn. Federal Judicial Center: Washington, DC.

Koehler, J.J. (1993) ‘Error and Exaggeration in the Presentation of DNA Evidence at Trial’ 34 *Jurimetrics Journal* 21.

- (2001) ‘The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient’ 74 *Southern California Law Review* 1275.

Koehler, J.J. and Shaviro, D.N. (1990) ‘Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods’ 75 *Cornell Law Review* 247.

Lempert, R.O. (1977) ‘Analyzing Relevance’ 75 *Michigan Law Review* 1021.

- (1991) ‘Some Caveats Concerning DNA as Criminal Identification Evidence: with Thanks to the Reverend Bayes’ 13 *Cardozo Law Review* 303.

- (1993) ‘The Suspect Population and DNA Identification’ 34 *Jurimetrics Journal* 1.

Lindley, D.V. (1991) ‘Probability’ in C.G.G Aitken and D.A. Stoney (eds.) *The Use of Statistics in Forensic Science*. Chichester: Ellis Horwood.

Redmayne, M. (1997) 'Presenting Probabilities in Court: The DNA Experience' 1 *International Journal of Evidence & Proof* 187.

- (1998) 'Bayesianism and Proof' in M. Freeman and H. Reece (eds.), *Science in Court*. Aldershot: Ashgate.
- (2001) *Expert Evidence and Criminal Justice*. Oxford: OUP.
- (2008) 'Exploring the Proof Paradoxes' 14 *Legal Theory* 281.

Roberts, P. (2009) 'The Science of Proof: Forensic Science Evidence in English Criminal Trials' in J. Fraser and R. Williams (eds.), *Handbook of Forensic Science*. Cullompton, Devon: Willan Publishing.

Roberts, P. and Zuckerman, A. (2010) *Criminal Evidence*, 2nd edn. Oxford: OUP.

Robertson, B. and Vignaux, G.A. (1992) 'Unhelpful Evidence in Paternity Cases' 9 *New Zealand Law Journal* 315.

- (1993) 'Probability - The Logic of the Law' 13 *Oxford Journal of Legal Studies* 457.
- (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom.* Chichester: Wiley.

Saks, M.J. and Koehler, J.J. (2005) 'The Coming Paradigm Shift in Forensic Identification Science' 309 *Science* 892.

Schum, D.A. (1994) *The Evidential Foundations of Probabilistic Reasoning.* Evanston, Illinois: Northwestern UP.

Stoney, D.A. (1991) 'What Made Us Think We Could Individualise Using Statistics?' 31 *Journal of the Forensic Science Society* 197.

Thompson, W.C. and Schumann, E.L. (1987) 'Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defence Attorney's Fallacy' 11 *Law and Human Behaviour* 167.

Thompson, W.C, Taroni, F., and Aitken, C.G.G. (2003) 'How the Probability of a False Positive Affects the Value of DNA Evidence' 48 *Journal of Forensic Sciences* 47.

Tribe, L.H. (1971) 'Trial by Mathematics: Precision and Ritual in the Legal Process' 84 *Harvard Law Review* 1329.

Usher, M.A. and Stapleton, R.R. (1979) 'An Unusual Forensic Application of Blood Group Studies' 19 *Medicine, Science and the Law* 165.

Zeisel, H. and Kaye, D. (1997) *Prove it with Figures: Empirical Methods in Law and Litigation*. New York: Springer.

BASIC PROBABILITY EXERCISES

Note to instructors: The goal of the exercises below is to help students understand (a) why probabilistic reporting is important and (b) how to calculate and understand simple probabilities. They start with a simpler scenario than a fingerprint, an analogy that should be closer to students' experience, and increase in complexity from there. Administer them by whatever means are appropriate for the class—written or verbal, as a class, in smaller groups, or individually.

Problem 1:

A robbery at night ends with the culprit escaping in a getaway car. A witness didn't get much of a look (it was dark and happened fast), but saw clearly that the first digit of the license plate was an X. Investigators have identified a car, and the first digit of the license plate is X: all of the information from the witness matches. How helpful is this "match" between plates? If you were on a jury, would you be likely to convict the driver of the car based on this "match" alone?

> [**A:** Hopefully not! You have no idea how many other cars might have that starting letter, but it's hardly definitive evidence that the driver committed the robbery.]

Problem 2:

A new witness comes forward in the same robbery, and says she saw the *last* three digits of the plate: 64T. She also noticed the car was a convertible with the top up. Investigators have found another car, a convertible with a license plate XW964T. This convertible is a "match" to all information we have from both witnesses. Is this "match" more, less, or equally helpful to the case than the "match" in Problem 1? *Why*?

> [**A:** As is hopefully obvious, this is much stronger evidence. In Problem 1, there was a "perfect" match between what the witness saw at the scene and the suspect car: both were cars, and both had a license plate starting with X. In Problem 2, there was again a perfect match. The difference is that in Problem 1, there are likely many cars that match the description equally well, because the description contained very little information. In Problem 2, with more information from witnesses, we expect fewer cars in town to have four specific license plate digits and be convertibles. This makes the evidence stronger, by narrowing the pool of suspects.]

Problem 3:

In the above scenarios, a witness was able to describe part of the license plate of the getaway car, but not all, and the suspect car matched that information. If you were a judge or juror, would it be enough for an expert witness to tell you "the car was a match"? Under what circumstances would you accept this simple explanation of the evidence?

> [**A:** With license plates, it's easy to draw the line at which it would be legitimate to tell a fact finder there was a match without elaborating: when the entire license plate matches, including the state. This is because license plates are unique *by design*, and its uniqueness can be

confirmed through DMV records. In other words, we know the absolute frequency of any particular combination of license plate number and state: 1. So, assuming the license plate was *real and belonged to the car it was seen on*, a full-plate match means that it was the same car. The expert witness should include the degree of uncertainty in any testimony, especially if the license plate is only partial.]

Problem 4:

How can a witness understand—and report to the fact finder—the strength of the evidence, the level of uncertainty about whether the cars match? Understanding probability can help here.
Start with expected frequency. To find out the frequency of license plates starting with X, you need to have some empirical knowledge of that frequency in the population in question, or you need to make some assumptions. If you have no access to DMV records, assume that license plates are assigned completely randomly, and that the first digit could be any letter or number. What portion of license plates would you expect to start with X?

[**A:** Assuming random assignment, there are 26 letters in the alphabet (A–Z) and 10 numerical digits (0–9), so 1/36 license plates would be expected to start with any one digit, X included.]

Problem 5:

A new witness came forward! This witness saw the first *two* digits of the license plate. She confirmed the first witness's account of it beginning with an X, and added that the second digit was 4. What's the expected probability of a license plate starting with X4?

[**A:** Since we're assuming license plates are randomly generated, the first and second digit should not affect each other's probability, so we can use the product rule: 1/36 *1/36 = 1/1296.]

Problem 6:

Let's say that the DMV databased tells us that only 1/60 license plates start with X. Knowing the expected probability of X license plates in the population in general, can you say anything about how likely the driver of the car the investigation found is the robber? Put another way, what is the probability that a driver of a car starting with X is innocent?

[**A:** If you answered "59/60 = .98, so I'm 98% certain he did it," or "he has a 1/60 chance of being innocent," then congratulations, you've committed the prosecutor's fallacy. 1/60 is the probability of any plate starting with X. Phrased as a conditional probability, that's the probability of a plate starting with X *if* the person's innocent. It's easy for someone to interpret this as the probability of a person being innocent *if* it starts with X, especially since that's really what we want to know.

Look at it this way: if the robbery took place in a town with 100,000 people with cars, and 1/60 of them are expected to have plates starting with X, 1,666 cars in that town are expected to have that license plate pattern. That means that all we've done is narrow the suspect pool to about 1,666 people, which means any one of those people has a 1/1,666 chance of being the robber—a very different probability from 59/60.]

Problem 7:

Many expert witnesses describe evidence in the form of a likelihood ratio. Calculate a likelihood ratio using the information in Problem 6.

> [**A:** The likelihood ratio is P(Evidence|Same Car)/P(Evidence|Different Car), or the probability of the license plate beginning with X if it's the car from the crime scene (a conditional probability) divided by the probability of the license plate beginning with X if it's not the car from the crime scene. Assuming the witness was correct, the first probability is 1. The second probability is the probability of any other car having a license plate starting with X, or 1/60. 1/(1/60) = 60.]

Problem 8:

Whatever form an expert witness's report takes, whether as a likelihood ratio, a simple frequency in the population, or any other description of the strength of a match, the point is to account for uncertainty. The fact finder should understand that a "match" does not guarantee that the known (suspect) and unknown (crime scene) items come from the same source (i.e., are the same car), because there may be duplication, and these expressions of probability help explain that uncertainty. But there are other possible reasons for a reported match between different cars, besides duplication, that are not accounted for in that expression. Name some other sources of uncertainty.

> [**A:** Other sources of uncertainty include: witness error, witness deception, someone else was driving the car, license plate faked or swapped out, and so on. Students should understand that the reported probability does not account for all types of uncertainty.]

Scientific Working Group on Friction Ridge Analysis Study and Technology, “Document # 21: Standards for Conclusions,” (Draft for Comment) (2003), 1-5. Document in public domain.

SWGFAST

Scientific Working Group on Friction Ridge Analysis, Study and Technology

Document #21

Standard for Consultation (Latent/Tenprint)

DRAFT FOR COMMENT

1. Preamble

1.1. As in any scientific endeavor, an examiner may have a need to discuss the examination with another analyst. This is generally referred to as a “consultation”.

1.2. Consultations are a natural and positive part of the scientific process. Consultations should be supported as part of the process and must be documented. Consultations may occur at any stage of analysis, comparison, evaluation, and evaluation (ACE-V), both before and after decisions are made during the examination. Consultations may result in recognition of differences of determinations or conclusions, creating "conflicts"[1]. Consultations are also used as part of the process to address conflicts, in accordance with quality assurance policies and in lieu of more formal conflict resolution procedures.

2. Scope

This standard provides guidance for examiners when consulting with one another during ACE-V examinations and related documentation requirements. This standard also provides guidance and model examples of how consultation can be used to resolve differences of opinion.

3. Definition of Consultation

3.1. A consultation [1] is a significant interaction between examiners regarding one or more impressions in question.

3.2. An interaction is considered significant when the consultant examiner (hereafter “consultant”) conducts an Analysis or Comparison of the impression(s). Specific examples are given in Section 4. The commonality among these examples is that they include, at a minimum, an Analysis of the impression(s), and may also include a comparison and evaluation.

3.3. Examples are given in Section 4 of discussions falling below the level of a significant interaction that typically involve minimal (or no) analysis. These typically have less potential to impact the key decision stages of ACE-V and are often related to case efficiency, strategy for workflow, or case management.

4. Examples of Significant and Non-Significant Interactions

4.1. Significant interactions rising to the level of consultation

4.1.1. Specific examples of discussion between examiners that are significant enough to rise to the level of consultation include the following:

4.1.1.1. Value determinations in analysis

4.1.1.2. Presence of significant distortions impacting the analysis or comparison

4.1.1.3. Presence of specific features during the analysis or comparison

4.1.1.4. Simultaneity of impressions

4.1.1.5. Whether an examination is complex or non-complex [2]

4.1.2. The reasoning for including the previous categories is because all of these consultations include at a minimum an analysis of the impression(s) and may also include a comparison and evaluation.

4.2. Interactions that are discussions, not rising to the level of consultation

4.2.1. Examples of discussions between examiners that do not rise to the level of significance to be designated as consultations include the following:

4.2.1.1. Suitability for automation fingerprint identification system (AFIS) entry

4.2.1.2. AFIS parameters

4.2.1.3. Administrative decisions such as triage

4.2.1.4. Searching efficiency “search smart clues”

4.2.1.5. Processing choices

4.2.1.6. Anatomical origin

4.2.1.7. Orientation

4.2.2. These discussions typically involve minimal (or no) analysis. They typically have less potential to impact the key decision stages of ACE-V and are often related to case efficiency, strategy for workflow, or case management.

4.3. There may be situations where a discussion rises to the threshold of a consultation because it has a significant impact on the case.

4.4. If there is doubt whether a discussion has risen to the level of a consultation, it should be treated as a consultation.

5. Documentation of a Consultation

5.1. The purpose of documenting a consultation is to record information or guidance obtained as a result of the consultation [3].

5.2. Consultations must be documented in the case record (e.g., analyst bench notes, a laboratory information management system). Discussions or other communications that do not reach the level of a consultation do not need to be documented.

5.3. The documentation for a consultation must include the following [3]:

5.3.1. Specific friction ridge impression(s) reviewed

5.3.2. The nature and result of the consultation

5.3.3. Initials, signature, or equivalent (e.g., unique identifier for the examiners involved)

5.3.4. Date

5.4. Depending on the nature and extent of the consultation, the consultant examiner may satisfy the above minimum documentation requirements by including the information within the notes of the initial analyst. It is also possible in more extensive consultations that a separate set of notes, annotations, or the consultant may generate images. These must be included in the case record. Examples are provided in Section 7.

6. Consultants for Comparison or Evaluation Should Not Subsequently be Used as Verifiers

6.1. An examiner who acts as a consultant during the comparison or evaluation of an impression shall not be used as the verifier for that impression.

6.2. An examiner who acts as a consultant during the analysis phase can be used as the verifier for that impression.

7. Case Examples

7.1. Example 1: Simple case with agreement between examiners

7.1.1. Consultation scenario

7.1.1.1. Examiner A is reviewing friction ridge impressions on a glass bottle received as an item of evidence (exhibit). Examiner A is unsure if one impression in particular has value for individualization, or should be considered of value for exclusion purposes only. Examiner A consults with Examiner B (the consultant examiner). Examiner B analyzes the impression on the bottle and determines that, in his opinion, it is "of value for individualization". Examiner B shares his determination with Examiner A. Examiner A agrees and declares the impression to be "of value for individualization".

7.1.2. Documentation

7.1.2.1. Examiner A's decision is recorded in his bench notes.

7.1.2.2. Examiner B initials next to the statement of value in Examiner A's notes. A date is included and the following phrase added to the margin of the notes: "Examiner B was consulted re: 'value' of latent print 1 (LP-1)". Examiner B's initials next to the decision of Examiner A indicates both the result of the consultation and that the examiners are in agreement.

7.1.3. Subsequent use of the consultant in Example 1 as a verifier

7.1.3.1. The consultant was used only during the analysis phase of the examination, not during comparison or evaluation. This examiner could be used as the verifier for the impression.

7.2. Example 2: Simple case with disagreement between examiners

7.2.1. Consultation scenario

7.2.1.1. Given the same scenario as in Example 1, Examiner B decides that LP-1 is 'of no value'. Examiner A originally had reservations about the 'of value for individualization' of LP-1, but has ultimately decided that LP-1 is 'of value for individualization'. A conflict has arisen, and has not been addressed by consultation.

7.2.2. Documentation

7.2.2.1. Examiner B initials next to the statement of value in Examiner A's notes. A date is included and the following phrase added to the margin of the notes: "Examiner B was consulted re: 'value' of LP-1. Examiner B said LP-1 is 'no value' because there are too few available characteristics."

7.2.3. Conflict resolution result

7.2.3.1. A conflict has arisen because there is a difference of opinion regarding a reportable conclusion. Consultation has not resolved the conflict and conflict resolution must now be initiated to resolve the disagreement [2].

7.3. Example 3: Complex case with agreement between examiners

7.3.1. Consultation scenario

7.3.1.1. Given the evidence as in Example 1, Examiner A has decided LP-1 is 'of value for individualization', has compared LP-1 against a suspect in the case, and is now struggling if he has 'sufficient agreement' between LP-1 and the suspect's right thumb exemplar to declare an 'individualization'. Examiner A approaches Examiner B and asks Examiner B to perform ACE examination of LP-1. Examiner B does so and decides that LP-1 can be individualized to the suspect's right thumbprint. Examiner B shares this decision with Examiner A. Both examiners sit down and review the examination together; discussing and conferring regarding which features were selected and compared to reach a decision. Specific challenges of the comparison are discussed and resolved. In the end, both examiners agree that LP-1 has sufficient agreement to report an 'individualization'.

7.3.2. Documentation

7.3.2.1. During the course of examination, both analysts generated bench notes with observations of LP-1. Both examiners generated annotated images, including an annotated Analysis image and an annotated Comparison image. All of these images are included in the case record. Examiner B includes his initials and date in the margins of the bench notes of Examiner A, next to the statement, "LP-1 individualized to Doe #1". In the margins, Examiner A writes "Consulted Examiner B re: sufficiency to ID LP-1; see additional notes."

7.3.3. Subsequent use of the consultant in Example 3 as a verifier

7.3.3.1. The consultant was used during the comparison or evaluation phases of the examination; therefore, the consultant could not be used as the verifier for the impression.

8. References

[1] SWGFAST Document #19 (2012), Standard Terminology of Friction Ridge Examination (Latent/Tenprint), Ver. 4.0.

[2] SWGFAST Document #10 (2013), Standard for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint), Ver. 2.0.

[3] SWGFAST Document #8 (2012), Standard for the Documentation of Analysis, Comparison, Evaluation, and Verification (ACE-V) (Latent), Ver. 2.0.

9. Revision Table

Version	Effective Start	Effective End	Posted	Archived	Change
1.0	3/14/13	N/A	4/27/13	N/A	Original Issue – Draft for Comment

CLASS 4

A. Biedermann, S. Bozza, F. Taroni, "Decision theoretic properties of forensic: underlying logic and argumentative implications" Forensic Science International, Vol. 177 (20 May 2008), 120-132 doi:10.1016/j.forsciint.2007.11.008.

Final draft post-refereeing
Final version published as:
Decision theoretic properties of forensic identification: Underlying logic and argumentative implications
A. Biedermann, S. Bozza, F. Taroni
Forensic Science International, Volume 177, Issues 2–3, 20 May 2008, Pages 120-132
doi:10.1016/j.forsciint.2007.11.008

Decision theoretic properties of forensic identification: underlying logic and argumentative implications

A. Biedermann [a,*], S. Bozza [b], F. Taroni [a]

[a] *The University of Lausanne, Ecole des Sciences Criminelles, Institut de Police Scientifique, le Batochime, 1015 Lausanne-Dorigny, Switzerland*

[b] *University Ca' Foscari, Department of Statistics, Venice, Italy*

Abstract

The field of forensic science has profited from recent advances in the elicitation of various kinds probabilistic data. These provide the basis for implementing probabilistic inference procedures (e.g., in terms of likelihood ratios) that address the task of discriminating among competing target propositions. There is ongoing discussion, however, whether forensic identification, that is, a conclusion that associates a potential source (such as an individual or object) with a given item of scientific evidence (e.g., a biological stain or a tool mark), can, if ever, be based on purely probabilistic argument. With regard to this issue, the present paper proposes to analyze the process of forensic identification from a decision theoretic point of view. Existing probabilistic inference procedures are used therein as an integral part. The idea underlying the proposed analyses is that inference and decision are connected in the sense that the former is the point of departure for the latter. As such the approach forms a coordinated whole, that is a framework also known in the context as 'full Bayesian (decision) approach'. This study points out that, as a logical extension to purely probabilistic reasoning, a decision theoretic conceptualization of forensic identification allows the content and structure of arguments to be examined from a reasonably distinct perspective and common fallacious interpretations to be avoided.

Key words: Bayesian decision theory, forensic identification, likelihood ratio

* Corresponding author. Tel.: +41 21 692 46 00
Email addresses: alex.biedermann@unil.ch (A. Biedermann),
franco.taroni@unil.ch (F. Taroni), silvia.bozza@unive.it (S. Bozza) (F. Taroni).

17 November 2007

1 Introduction

It is currently not controversial that there is no positive solution by which particular propositions [1] could conclusively proven to be true. Notwithstanding, it is a widespread practice among parts of the forensic community to reach and defend conclusions that seem to suggest the contrary. Among the more prominent instances of this kind are so-called 'identifications', more correctly termed 'individualizations' [2], a type of conclusion by which a person or thing is specifically distinguished from all other persons or things of the same kind.

As a consequence of uncertainties that inevitably accompany any forensically relevant real-world event, most state of the art identification procedures involve some sort of mathematically acceptable and statistically rigorous probabilistic component [3]. The ensuing conclusions operate in a continuous range and actors that aim at formulating definite conclusions of 'identification' (or 'exclusion') are bound to take an inferential step based on assumptions that go beyond what is logically warranted by the underlying inference procedure (justification for this claim is given later in the text). Questions of whether such definite conclusions should be attempted, if it can be done, and by whom, is at the heart of ongoing controversies in forensic and legal areas at large. Without entering into further details of these discussions, it is solely noted at this point that it is precisely that aforementioned, ultimate inferential step which remains to be the least formally managed one. Actually, its nature is considered as obscure [4] and has distinguished commentators led to consider forensic identification as unscientific [5].

The present paper focuses on a particular aspect of that debate, that is the process of extracting a particular conclusion *once* that target propositions in an identification scenario have been re-evaluated by means of a probabilistic procedure (such as, for instance, a likelihood ratio) and based on some given evidence. The proposed analysis intentionally avoids to rely on both the generation and the use of experimental data of any kind. Instead, the aim is to examine the 'problem' of forensic identification from a decision theoretic perspective, that is, a development that regroups probability theory and utility

[1] A proposition is interpreted here as a statement or assertion that such-and-such is the case (e.g., an outcome or a state of nature). It is assumed that personal degrees of belief can be assigned to it (Section 2).

[2] In the remainder of this paper, the two terms will, loosely speaking, be used interchangeably despite the fact that the latter is technically more correct than the former. In fact, a proper term for the process of addressing the issue of whether or not a particular item came from a particular source is 'individualization' [1], but established forensic and judicial practices now commonly refer to it as 'identification'. Following Kingston [2], the notion of 'identification' is, however, more accurately to be considered as a determination of some set to which an object belongs or the determination as to whether an object belongs to a given set.

theory within a coordinated whole. The basic tenets of decision theory are outlined in Section 2.

No effort will be deployed in arguing that probabilities are an appropriate means for measuring personal uncertainties. This has at length been done in existing literature on the topic [6]. The leading idea underlying this paper is different. The aim is to admit a collection of basic concepts of decision theory that can be regarded as demanding a least concessional attitude on behalf of the reader in order to see what − hopefully beneficial − argumentative and methodological consequences can be extracted from such a starting point.

2 Concepts and notation

For the purpose of the current discussion, the reader is invited to consider a structure for the problem of interest in terms of three major constituents, in much the same way as it is described in standard literature on the topic of decision analysis [7, e.g.]. The principal concepts are as follows:

(1) At the heart of the target problem is a collection of n real-world states, alternatively termed 'events', about which one is uncertain. These states of nature will be written $\theta_1, \theta_2, ..., \theta_n$. It will be assumed that uncertainties about these states reflect personal degrees of belief that can be formally expressed in terms of numbers whose properties and manipulation agree with the laws of probability theory. The probabilities of states of nature will be written $Pr(\theta_1), Pr(\theta_2), ..., Pr(\theta_n)$. More formally, on may also say that it is possible to measure uncertainty over Ω (that is, the entirety of the nature $\theta_1, \theta_2, ..., \theta_n$) by means of a suitable probability distribution Pr.

(2) A second major aspect of the proposed analyses is a collection of decisions, that is, available actions, written $d_1, d_2, ..., d_m$, among which a reasoner must make a choice. The entries of the list of decision are, as are those of the list of events (above mentioned point (1)), exclusive and exhaustive. The space of decisions $d_1, d_2, ..., d_m$ is sometimes abbreviated by Δ.

(3) The choice of a particular decision d_i when θ_j turns out to be the true state of nature leads to a distinct consequence, denoted C_{ij}. There is a total of $m \times n$ consequences and, as a third element of the notational and conceptual apparatus invoked at this point, each of these consequences has an associated expression of desirability. That desirability, also called utility, is written $u(C_{ij})$ and will also be handled in a numerical form. The way in which numbers may actually be assigned does not need, however, immediate consideration. This subject will be treated as the discussion proceeds. For the time being, it is solely noted that such numbers will allow one to quantify preferences.

In summary, the stated concepts allow a person who is in charge of making a decision to define a problem of interest in terms of consequences that each of the possible lines of available action will have given each of a set of mutually exclusive events about which that individual is uncertain.

Although the discussion in this paper is one that refers to some extent to modes of reasoning about the truth or otherwise of propositions, the classical theory of statistical inference will not be considered here. The main reason for this, among others, is a fundamental incompatibility that stems for the fact that a given identification scenario (of a particular case at hand) cannot be meaningfully conceptualized as a instance of some sort of replicable sequence of trials under stable conditions [7,8].

3 A decision theoretic conceptualization of identification

3.1 Uncertainty in target propositions

Forensic scientists are commonly called upon to examine material collected at a crime scene. For example, they may focus on analytical parameters of a crime stain, generate measurements and compare these, if available, to those extracted from a sample obtained from a suspect. As a result, a series of similarities and differences is usually noted and, subsequently, the forensic scientist may be required to offer assistance in evaluating how such information is amenable to discriminate between propositions that are of interest to a particular client (e.g., a prosecutor). Such evaluations ought to be done, as it is now almost universally accepted in the judicial area, according to the laws of probability theory where Bayes' theorem provides for the re-evaluation of the probability of a proposition given newly acquired evidence [9,10, e.g.].

The present discussion will not focus on how probabilistic evidential assessments are actually made. It is solely assumed, as part of a general acceptance, that one can and should assure that they agree with probability theory. In particular, the discussion will assume that some not further specified individual − referred here to as the decision maker − will be about to maintain personal beliefs about at least a pair of propositions, formulated within a fairly general standard identification scenario.

It may be, however, that probabilistic reasoning about such target propositions is not that what is requested by the client, or the scientist may not wish to engage in analyses of this kind. On the contrary, it may well be that conclusions in terms of definite assertions of 'identification' or 'exclusion' are either explicitly demanded by a client or be part of a scientist's sole accepted practice. It is important to note, however, that this is an aspect which − in a strict sense − is not covered by a purely probabilistic framework, that is, further considerations are required. As a central topic of this paper, the analyses outlined hereafter will point out this issue in further detail.

As a starting point, let there be a pair of states of the world (in forensic contexts commonly referred to as source-level [11] propositions) defined as follows:

- θ_1: the suspect is at the origin of the crime mark (e.g., a fingermark);
- θ_2: someone else is at the origin of the crime mark.

Let $Pr(\theta_1)$ and $Pr(\theta_2)$ be a decision maker's personal probabilities for these propositions. This is written short for a more extensive notation that covers a conditioning by (non-scientific) circumstantial information I and, if applicable, a particular item of (scientific) evidence E (e.g., in terms of observed correspondences in DNA profiles). That extended notation would take the form of, for example, $Pr(\theta_1 \mid E, I)$ for a conditioning covering both E and I, or $Pr(\theta_1 \mid I)$ for a conditioning on I only. In forensic contexts, a conditioning by at least I is always assumed although it is often omitted from notation for sake of brevity. As is readily seen, no particular distinction is drawn here between the notions of 'prior' and 'posterior' probabilities. The reason for this is the following. With respect to $Pr(\theta_1 \mid I)$, for instance, $Pr(\theta_1 \mid E_1, I)$ may be thought of as the 'posterior' probability of θ_1 given the additional item of information E_1 (and I). But, with respect to $Pr(\theta_1 \mid E_1, E_2, I)$ – where E_2 denotes a further item of evidence – the term $Pr(\theta_1 \mid E_1, I)$ may be interpreted as a 'prior' probability. A concise summary of this view is due to Lindley [12, p. 301]: "Today's posterior is tomorrow's prior."

According to the kind of evidence, there may also be further matter for discussion. In the case of DNA, for instance, it may be an issue whether θ_2 should account for particular population characteristics and/or allow for close relatives of the suspect [13]. For the purpose of the current discussion, it is solely noted that such considerations can in principle be incorporated [14] and the scientists is assumed to do so according to the requirements of the case at hand.

Next, let our hypothetical decision-maker – who need not necessarily be the scientist – consider a pair of probabilities $Pr(\theta_1)$ and $Pr(\theta_2)$. The key question that the reader is now asked to consider is the following:

> On the basis of states of personal belief $Pr(\theta_1)$ and $Pr(\theta_2)$, how is the decision-maker to *decide* whether or not to 'call an identification', that is, to individualize the suspect as the source of the crime stain?

Whether this either is what a scientist truly intends to do or his client expects him to do, one can readily see that this involves more of a problem than just a probabilistic evaluation. Actually, the scientist is about to make a decision while the true state of affairs – the suspect being or not the source of the crime mark – is uncertain. Stated otherwise, the key matter that the scientist is concerned with is one of a decision based on a probabilistic assessment. That decision process involves not only a probability of the respective state of nature but also a consideration of the desirability of the outcome.

For the time being, the discussion here shall ignore the fact that individualization are much more common in such contexts as friction ridge mark evidence than in settings involving, for instance, DNA evidence. Besides, there are also court rulings according to which scientists were not to express individualizations (e.g., in *R. v. Doheny* or *R. v. Adams*). In view of this, the purpose here is to study both the structure of the problem and an ordered way of approaching it independently from the actor actually faced with the process of individualization.

3.2 Interpreting the scope of available conclusions – as decisions

The act or process of 'identifying' an individual as being the source of a crime mark can be interpreted as one particular course of action, or, alternatively stated, a decision. Let it be noted d_1, for example. Although alternative conclusions may vary somewhat with the area of expertise, for many practitioners, in particular fingerprint experts, the collection of statements may also cover 'inconclusive' and 'exclusion' [15]. In the following, decisions to render one of the latter two conclusions will be denoted d_2 and d_3, respectively.

In the context, it seems important to insist on the fundamentally different nature of decisions and propositions. A proposition, as it is understood in Section 3.1, assumes exactly one state. A binary proposition, for instance, is either true or false although, usually, it is not known with certainty which of the two states actually holds. Typically, it is not known whether or not the suspect is the source of the crime mark. A decision maker who addresses a source-level proposition thus engages in a risky argument.

In contrast to this, a decision assimilates, according to the interpretation proposed above, to an expert's opinion. An actor must consciously decide that a conclusion of, for instance, 'the suspect is the source of the crime mark' (that is, an individualization) is the state of nature that he actually believes to be true. But this particular decision does not make that state of nature any more probable or improbable. Stated otherwise, probabilities of states of nature are assumed to independent from decisions[3].

In particular, recipients of expert opinion should be proficient in distinguishing decisions (expert opinions) from propositions because equating by default the latter with the former may be, according to the setting, devastating. For example, following an expert's conclusion that individualizes the suspect as the source of the crime mark while that mark has, in reality, been left by some other person, amounts to a false identification. We take the liberty to call

[3] Notice that this does not hold for general decision problems. According to the problem at hand, it may be preferable to write $Pr(\theta_j \mid d_i)$ instead of $Pr(\theta_j)$ alone.

Table 1
Decision table for an identification problem with d_i, $i = 1, 2, 3$ denoting decisions, θ_j, $j = 1, 2$ denoting states of nature and C_{ij} denoting consequences.

	$\theta_{1:}$ suspect is donor	θ_2: some other person is donor
d_1: identification	C_{11}: correct identification	C_{12}: false identification
d_2: inconclusive	C_{21}: 'neutral'	C_{22}: 'neutral'
d_3: exclusion	C_{31}: false exclusion	C_{32}: correct exclusion

such an argument 'identification fallacy' and provide further discussion on it in Section 5.3.

For the time being and in view of the fact that decisions and propositions are distinct ingredients of a decision problem, attention should also be drawn to their combination, that is, consequences. This is considered in the next Section.

3.3 Consequences, preferences and utilities

In evaluating the decision d_i that might be adopted while the true state of affairs θ_j is unknown, one is required to study the consequences C_{ij} to which d_i and θ_j combine, along with their associated desirabilities $u(C_{ij})$ (Section 2).

The model for the currently discussed scenario lends itself to a series of readily framable consequences. Table 1 gives a summary of this. Clearly, a decision to individualize a suspect (d_1) if he is truly at the origin of the crime mark (θ_1) represents a correct identification (C_{11}). Conversely, that decision (d_1) is an erroneous identification (C_{12}) if the suspect were not at the origin of the crime mark (θ_2). By an analogous line of argument, excluding a suspect (d_3) who truly is at the origin of a crime mark (θ_1) is a false exclusion (C_{31}) whereas a correct exclusion (C_{32}) holds when the suspect were truly not at the origin of the crime mark (θ_2). The setting as considered so far is equivalent to the problem of correctly diagnosing, for instance, whether or not a patient is affected by a particular disease.

The decision to conclude 'inconclusive' (d_2) does not convey any information that tends to associate or otherwise the suspect with the source-issue. Therefore, Table 1 lists the consequences C_{21} and C_{22} as 'neutral'. The reader may wish to use his own term. The point solely is that the consequence, as such, does not seem to represent an immediate harm to the situation of the suspect.

As may be seen, there are consequences that are far more repercussive than others and this brings discussion to the assessment of their respective desirability (utility). Later in the text, it will be shown that utilities of consequences are needed, along with probabilities of obtaining the respective consequences, to determine the decision with the highest *expected utility* (there may be more than one such decision if they have the same expected utility). A more formal statement of the notion of 'expected utility' is given in Section 3.4 (Equation 2).

As noted previously in Section 2, consequences are defined in such a way that it is possible to rank them according to their desirability. For the various consequences set forth in Table 1, one may agree about the following ranking:

$$(C_{11}, C_{32}), (C_{21}, C_{22}), C_{31}, C_{12} \tag{1}$$

In this ranking, decreasing preferences are assumed from left to the right and equally preferred consequences are written between brackets. An important assumption underlying that ranking is its transitivity. That is to say, if, for instance, the consequence C_{21} is preferred to C_{31} and C_{31} is preferred to C_{12}, then C_{21} is preferred to C_{12}.

We insist on Equation 1 being interpreted as a suggested ranking. It is motivated in some detail below but the reader is free to consider his own ranking.

The ranking (1) reflects that a correct identification (C_{11}) and a correct exclusion (C_{32}), that is, accurate conclusions, are the most preferred consequences. At the other end of the scale is, as the least preferred consequence, a false identification (C_{12}). A false exclusion (C_{31}) may be considered less undesirable than a false identification (C_{12}) because it should not negatively affect the situation of the suspect. But yet, a false exclusion (C_{31}) is of undesirable character because it constitutes an erroneous conclusion that legal proceedings generally seek to avoid.

Decisions to state 'inconclusive' assume a somewhat intermediate status. They reflect no opinion about the true state of affairs (that is, the suspect's true connection with the crime mark). This emptiness of decision d_2 is in exchange to the absence of any 'risk' of rendering an inaccurate conclusion. So, a decision to render a noncommittal 'inconclusive' statement (d_2), independently of whether θ_1 or θ_2 (and consequently C_{21} or C_{22}) holds, is preferred to inaccurate conclusions (C_{31} and C_{12}) but less preferred to an informative, accurate identification (C_{11}) or exclusion (C_{32}).

So far, the suggested ranking (1) expresses preferences among consequences. The ranking encapsulates coherent comparisons, but as such, it is not of a more definite form. In order to achieve a more directly manageable form − one that will allow it to combine more easily with probabilities and decisions −, the desirability of consequences may be measured by a number called u, the utility (where $u \in (0, 1)$) [4]. The attractiveness of consequence C_{ij}, that is its utility, is thus written $u(C_{ij})$ or $u(d_i, \theta_j)$.

[4] Confining utility values to (0,1) is of some convenience for the purpose of the current discussion but we note that there are areas of application where other scalings may be applied (e.g., in economy) [16].

Table 2
Decision table with decisions d_i (d_1 : 'identification', d_2 : 'inconclusive', d_3 : 'exclusion'), states of states of nature θ_j (θ_1: the suspect is at the origin of the crime mark, θ_2 : some other person is at the origin of the crime mark), probabilities for states of nature $Pr(\theta_j)$, consequences C_{ij} and utilities $u(C_{ij})$.

	θ_1	θ_2
d_1	$u(C_{11}) = 1$	$u(C_{12}) = 0$
d_2	$u(C_{21}) = \alpha$	$u(C_{22}) = \alpha$
d_3	$u(C_{31}) = \beta$	$u(C_{32}) = 1$
Probabilities	$Pr(\theta_1)$	$Pr(\theta_2)$

Table 2 provides a summary of the various utilities that need to be thought about. Let us note that at this stage, attention is solely drawn on the rationale behind the $u(C_{ij})$ and not the numbers and variables to which these expressions are actually equated. We will follow the argument that the assignment of numerical values (confined between 0 and 1) to consequences can be made through a coherent comparison with a standard that is given by two reference consequences [7]. These may be chosen, respectively, as the best (written C) and the worst (written c) of all consequences listed in Table 2.

According to preceding discussion, the pair (C_{11}, C_{32}) (i.e., accurate conclusions) covers the most desirable consequences and their associated utilities are set to 1. A false identification (C_{32}), the least desirable consequence, is assigned a 0 utility.

The pair (C_{11}, C_{32}) and C_{32} represent, respectively, the most (C) and the least (c) desirable consequences. They are used as reference points for deriving utility values for the remaining consequences C_{ij}, each of which can be compared favorably with c and unfavorably with C.

The utilities of the consequences of 'intermediate' desirability can now be assessed as follows. The utility of a 'neutral' consequence (C_{21} or C_{22}) is the number, let it be called α, that would make one indifferent between the option 'C_{21} with certainty' and 'C with probability α' (i.e., 'c with probability $1-\alpha$'). Stated otherwise, the utility of C_{21} is the probability α with which the best consequence could be obtained and which would make that option just as desirable as the option of obtaining C_{21} with certainty (an analogous argument with probability $1-\alpha$ holds with respect to the worst consequence). Thus, considering C_{21} and C_{22} equally desirable (Ranking (1)), one has $u(C_{21}) = u(C_{22}) = \alpha$. Similarly, β is proposed as a value that would make one indifferent between the options 'C with probability β' and 'C_{32} with certainty'. An example of the application of such modes of analysis is presented in [17].

In our running example, one can reasonably expect to find agreement on values that satisfy $\alpha > \beta$. Notwithstanding, it is generally advisable to guard against inconsistencies by checking actual values against other combinations of reference points [16, e.g.]. Further details on this topic are provided in Appendix A.

The discussion of specific numerical values is delayed until later. It will be argued that these utilities are intimately related to the decision maker and that they will have a crucial bearing in evaluating the decision that can be expected to lead − given a particular $Pr(\theta)$ − to an optimal consequence. This latter topic is now examined in more detail.

3.4 Expected merit of the various courses of action

The essential ingredients of a typical identification scenario have now been framed. The scope of possible conclusions, interpreted as decisions, define, along with the current true state of affairs, a collection of consequences. At the end of the previous Section, discussion has focused on how the desirability of each of these consequences may be elicited. But all of these assessments may not, however, be of immediate help to the decision maker. The various elements of the problem description need to be combined in order to obtain practically useful indications.

For example, utilities have been assessed for consequences that are the result of a decision d_i, $i = 1, ..., m$, when θ_j, $j = 1, ..., m$, turns out the be true state of nature. The actual situation faced by a decision maker is different, however. He may wish to learn about the expected utility of a decision d_i while being uncertain about the actual state of nature. There is an expression for this, known as 'expected utility of a decision d_i', written $\bar{u}(d_i)$, that is obtained by weighting the possible consequences of that decision by the probability with which they may be obtained. Generally, using habitual notation introduced so far, the expected utility of a decision d_i is:

$$\bar{u}(d_i) = \sum_{j=1}^{n} u(C_{ij}) Pr(\theta_j). \tag{2}$$

Applied to the problem under study, the expected utility of a decision to individualize (d_1) is $\bar{u}(d_1) = u(C_{11})Pr(\theta_1) + u(C_{12})Pr(\theta_2)$. Following the utilities defined and summarized in Table 2, it is readily seen that the expected utility of decision d_1 reduces to:

$$\bar{u}(d_1) = Pr(\theta_1) \tag{3}$$

Analogously and referring again to Table 2, the expected utilities of the decisions 'inconclusive' (d_2) and 'exclusion' (d_3) can be found as follows:

$$\bar{u}(d_2) = u(C_{21})Pr(\theta_1) + u(C_{22})(1 - Pr(\theta_1))$$

$$
\begin{aligned}
&= \alpha Pr(\theta_1) + \alpha(1 - Pr(\theta_1)) \\
&= \alpha
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
\bar{u}(d_3) &= u(C_{31})Pr(\theta_1) + u(C_{32})(1 - Pr(\theta_1)) \\
&= \beta Pr(\theta_1) + (1 - Pr(\theta_1))
\end{aligned} \tag{5}
$$

There is a noteworthy situation in which a symmetrical relation appears between this latter relation and the expected utility of an identification. That is the case when a false exclusion (C_{31}) is judged just as undesirable as a false identification (i.e., by assigning it a zero utility) while leaving the utilities of accurate identifications (C_{11}) and exclusions (C_{32}) as defined above (i.e., utility of 1). In such a setting $\bar{u}(d_3)$ reduces to $1 - Pr(\theta)$ and thus is the complement of $\bar{u}(d_1)$.

Besides the expected utility of each decision, another potentially relevant question may relate to the decision that the decision maker finally ought to adopt. If the aim is to act in such a way that the expected consequences are optimal with respect to their desirability, then the procedure would consist of choosing that decision which has the maximum expected utility, written more formally as:

$$
\max_i \bar{u}(d_i) \tag{6}
$$

For the currently studied problem, there are, in summary, the following three expected utilities:

$$
\bar{u}(d_1) = Pr(\theta_1),\ \bar{u}(d_2) = \alpha,\ \bar{u}(d_3) = \beta Pr(\theta_1) + (1 - Pr(\theta_1)).
$$

It cannot be told immediately which of these expressions maximizes expected utility as long as nothing is said about the relative magnitude of $Pr(\theta_1)$, α and β.

An important aspect of this intermediate conclusion is that the proposed decision model requires more than just a consideration of the probability of the truth or otherwise of a target proposition (e.g., 'the suspect is the source of the crime stain'). A comparison is needed among alternative courses of action, notably in terms of the component factors from which one determines the respective desirabilities of expected consequences.

4 Investigating the decision model

4.1 Maximizing expected utility

The aim of investigating the proposed model is to learn more about the characteristics of a situation, given one's personal preferences, in which opting for one or another decision is advisable. The approach outlined in Section 3 involves placeholders for numbers that assume values from a distinct range. In principle, the procedure can be worked through using some given numerical values, but one might be reluctant to do so because one might want gained insight to be independent from a specific body of numerical assessments.

A feasible attempt to remain essentially non-committal with respect to particular numerical values consists of comparing the various functions by plotting them against each other. This allows certain variables to be detached from specific numbers. Such a plot is given in Figure A.1. This Figure represents Equations (3), (4) and (5) that express the expected utility of the decisions d_i $(i = 1, 2, 3)$ as a function of $Pr(\theta_1)$. For the purpose of illustration, the values 0.8 and 0.2 have been assigned to, respectively, α and β.

When the aim is to determine the decision that maximizes expected utility − given a particular value for $Pr(\theta_1)$ − then the task consists of focusing on that decision (d_i) whose expected utility ($\bar{u}(d_i)$) is greater (or, at least not lower) than that of any other decision. For any value $Pr(\theta_1)$, the graph in Figure A.1 is readily inspected to find the decision that maximizes expected utility. Actually, by following the lines one can also read off the graph the points at which the values of one function become to lie above those of other functions. Figure A.1 illustrates this with a bold line. That line marks what is sometimes referred to as the 'Bayes decision' or 'Bayes action', that is, the decision that maximizes expected utility.

The Bayesian action depends on the actual values assigned to α, β and on the probability $Pr(\theta_1)$. It is clear that different values will lead to scenarios that differ from that illustrated in Figure A.1. Actually, one can face the following two main sets of cases:

- *Case 1*: Each decision can be, at least for some values of the probability of θ_1, the Bayes action.
- *Case 2*: The set of decisions that maximize expected utility is restricted to d_1 and d_3.

Some major characteristics of these two categories of cases are discussed below.

4.1.1 Case 1

Figure A.1 shows an example for this category of cases. Scenarios of this kind arise whenever the utility values α and β are such that $\alpha > \frac{1}{2-\beta}$, where $\frac{1}{2-\beta}$

is the x-coordinate of the point of intersection between $\bar{u}(d_1)$ and $\bar{u}(d_3)$. One may then consider, for example, the effect of assigning higher utility values α to neutral consequences. For a fixed value β and for any value of α such that $\alpha > \frac{1}{2-\beta}$, the point of intersection between $\bar{u}(d_2)$ and, for example, $\bar{u}(d_1)$ is shifted towards higher values of $Pr(\theta_1)$. An increased utility of a 'neutral' consequence thus means that one requires a higher prospect of obtaining the best consequence (correct identification, C_{11}), that is, a prospect that would make one indifferent to the option of obtaining a 'neutral' result with certainty. Alternatively, one may also say that an increased expected utility of $\bar{u}(d_2)$ tends to narrow the range of values $Pr(\theta_1)$ (towards the upper end of the interval between 0 and 1) for which d_1 maximizes expected utility.

Yet another way to express the effect of assigning higher utilities α to neutral consequences would be to say that the decision d_1 becomes preferable (not generally, but compared to d_2) for a smaller range of probability values $Pr(\theta_1)$: by increasing α there will be more distinct cases (that is, cases with differing $Pr(\theta_1)$) in which the probability of the suspect being the offender (that is, the probabilities for obtaining a correct identification if decision d_1 were chosen) is below that value α.

Notice that these statements analogously apply with respect to the intersection between $\bar{u}(d_2)$ and $\bar{u}(d_3)$. In summary, thus, assigning greater preference to 'neutral' consequences means to narrow the range of probability values for which decisions d_1 and d_3 lead to, respectively, false identifications and exclusions. These all seem to be entirely reasonable properties.

These thoughts on the value α involve more, however, than a mere manipulation of variables and equations. It is important to notice that the proposed approach actually allows one to capture and study a real situation faced by an individual − that of the decision maker. For the purpose of illustration and by referring again to α, contemplating this value could be assisted by questioning oneself as follows:

> There are values for $Pr(\theta_1)$ so that if one were to select, for example, decision d_1, the prospect of obtaining a correct identification is not high enough to make one indifferent to the option of an 'inconclusive'-decision from which one were sure that it would not result in any erroneous identifications or exclusion. These values for $Pr(\theta_1)$ are such that they are smaller than α.

4.1.2 Case 2

When the utility α of the consequences C_{21} and C_{22} is smaller than $\frac{1}{2-\beta}$, then the set of decisions that maximize expected utility is confined to d_1 and d_3 (e.g., Figure A.2). A decision maker with such a preferential setting will actually concentrate his decision making to identifications and exclusions. Such a position is interesting to investigate in further detail because, on a rough 'intuitive' view, such 'bold' and definite conclusions are frequently said, from a client's point of view, to be most useful.

Admittedly, the idea of a preferential setting that leads to a decision policy that covers only conclusions of identification or exclusion is somewhat hypothetically extreme, but it serves the purpose of clarifying that tuning one screw of the model may well enhance one aspect, but possibly to the detriment of other aspects.

Consider this in a setting mentioned earlier in Section 3.4, where the expected utilities were $\bar{u}(d_1) = Pr(\theta_1)$ and $\bar{u}(d_3) = 1 - Pr(\theta_1)$ (that is, a limiting case where $u(C_{31}) = \beta = 0$). Here, the expected utilities of d_1 and d_3 coincide if $Pr(\theta_1) = 0.5$.

It may be worthwhile to enquire about whether this setting has convenient implications. At first glance, the binary decision-policy seems to work well because whenever the probability of the suspect's being the source of the crime mark ($Pr(\theta_1)$) is close to either 1 or 0, then choosing, respectively, decision d_1 or decision d_3 will lead to a correct conclusion with a high probability.

One must be aware, however, that a strict identification-exclusion policy needs to be thought through for whatever value $Pr(\theta_1)$ assumes and this is where the policy might exhibit undesirable properties. In particular, it may not always be the case that $Pr(\theta_1)$ is close to 1 or 0. There is a whole range of intermediate values $Pr(\theta_1)$ and if for these too only d_1 or d_3 maximize expected utility, then one cannot avoid accepting erroneous conclusions to occur with increased probability. The situation is most acute if $Pr(\theta_1)$ is 0.5. Then, decisions d_1 and d_3 have the same expected utility and choosing one of them will lead with a probability of 0.5 to an erroneous conclusion (that is, a false identification if d_1 is selected and a false exclusion if d_3 is selected).

4.2 Qualitative evaluation of preferential relationships

The currently studied decision theoretic approach captures forensic identification in terms of consequences that are the result of a combination between particular decisions (that is, statements of the kind 'identification', 'inconclusive' and 'exclusion') and a given state of affairs (the suspect's being or not the source of the crime mark). Complication stems from the fact that the true state of nature is unknown so that it cannot readily be determined which of the available decisions would lead to the most optimal consequence.

Given the crucial role of both decisions and states of nature, one may be required to articulate and inform how these ingredients influence the decision analysis process. Again, insight in the model's underlying properties may be most valuable if they can be formulated independently from specific numerical assessments. Below, this is considered in some detail.

As a starting point, Figure A.3 is proposed with the aim of clarifying the formal relationship between the various components. Actually, this graph is an influence diagram, that is, a Bayesian network extended by a decision and a utility node (represented as, respectively, a square and a diamond). The interested reader can use conventional Bayesian network software to implement this model. The definition of the decisions (node D), the probabilities for the states of nature (node θ) and the utilities of the consequences (node U) follows the specification given in Table 2.

Figure A.3 reflects the dependence of utilities on decisions and states of nature in terms of directed edges that emanate from both nodes θ and D. These two influences can be examined separately. For example, if one were to consider the effect that changes in the state of nature have on expected utility, one needs to compare the utilities $u(C_{i1})$ and $u(C_{i2})$ for each decision d_i. Following the preferences set forth in Table 2 it can be seen that a change in the truthstate of θ has varying effects on the utility of the respective consequences: for d_1 $u(C_{11}) > u(C_{12})$, for d_2 $u(C_{21}) = u(C_{22})$ and for d_3 $u(C_{31}) < u(C_{32})$. Thus, one cannot tell how a change in the state of affairs (θ) affects expected utility without making reference to a particular decision.

A similar observation holds for the effect due to a change from one decision to another. Here, one needs to compare the utilities $u(C_{1j})$, $u(C_{2j})$ and $u(C_{3j})$ for each state of nature θ_j. For this comparison, consider

$$d_1, d_2, d_3$$

as an ordering of the target decisions with decreasing 'associative effect' (for the suspect with respect to the crime mark) from left to the right. Now, if the suspect is truly the source of the crime mark (θ_1), then the utilities compare as follows (Table 2): $u(C_{11}) > u(C_{21}) > u(C_{31})$. However, if someone else is the source of the crime mark (θ_2), then the utilities of the respective consequence compare in exactly inverse fashion: $u(C_{12}) < u(C_{22}) < u(C_{32})$.

What can be retained from this comparison is that one cannot generally tell − independently from the current state of affairs (θ) − how the change from one decision to another would affect expected utility. What can be said is, however, that for someone whose set of preferences is as defined in Table 2, the expected utility of a decision is the higher the more 'associative' the decision is and if θ_1 holds. The tendency of the expected utility behaves analogously for less 'associative' decisions and if θ_2 holds.

4.3 Decision analysis: a logical extension to probability calculus

The proposed decision theoretic analyses draw a clear distinction between the probability of a proposition and the decision that an individual may take with respect to that proposition. It should be emphasized that this conceptualization of forensic identification − as a decision − is not detached from the purely probabilistic apparatus that is traditionally applied for evaluating uncertainties about the truthstate of propositions. On the contrary, the two apparatuses are closely related and the influence diagram considered in the previous Section (Figure A.3) allows to clarify this point.

θ is an uncertain proposition and in Section 3.1 it has been said that it is always conditioned upon other circumstantial information I along with, if applicable, some given evidence E. That conditioning is obtained through application of Bayes' theorem, a kind of computation that can be supported by graphical probability models, notably Bayesian networks − even if the ensuing probabilistic calculations may be of increased complexity [18]. In particular, one can have a Bayesian network where the proposition θ ('the suspect (some other person) is the source of the crime mark') is a node that conditions particular scientific evidence (e.g., an observed correspondence between a crime mark and a reference print). In the area of DNA evidence evaluation, for example, this is currently possible through complex and well developed models [19]. In turn, a node θ can also serve as a basis for reasoning about a crime-level proposition (e.g., 'the suspect (some other person) is the offender') or about evidential relevance [20].

The reason for this short extension of argument is to point out that a proposition such as θ, which is part of the proposed decision model (Figure A.3), can be viewed as an instance of another (possibly larger) inference network (see for example [21]). This means that, in the course of a decision analysis, uncertainty in relation with θ can be informed coherently in terms of a probabilistic evaluation. Thus, the two tasks are logically interrelated and operational means (e.g., Bayesian networks) exist to assure that the two tasks neatly interface. Decision analysis is not just some additional concept among others, but an integral part of a coherent approach to practical action for which probability is merely an essential preliminary.

5 Discussion

5.1 Identification/individualization: a case of problematic suppression of uncertainty

The discussion has been given different labels and is known as the problem of 'inference to common source', 'source attribution' or 'individualization/identification [5] '. The scope of this paper does not allow to give a comprehensive account of all facets of this subtile subject that has been substantially influenced by writers such as − to mention only some of the more recent

[5] The distinction between 'individualization' and 'identification' is briefly mentioned in Section 1.

and widely referenced ones – Champod [3,4], Evett et al. [22], Inman and Rudin [23] as well as Stoney [24].

The discourses essentially gravitate around the question of how, if ever, individualizations can be reached on the basis of probabilistic considerations. To date, no unanimous answer appears to exist. As a matter of fact, individualizations are reached in current practice but it is held that such conclusions cannot, from an argumentative point of view, derive from any scientific process [5]. The scope of diverging opinions is large indeed as is illustrated by the fact that some writers argue contrarily and even suggest that the study of the structure of arguments (e.g., based on models for statistical treatment) is not applicable to all forensic disciplines [25, e.g.].

Anybody who engages in a thought process that aims at individualizing a single individual or object as the source of a mark or stain believed to be relevant to a case will be about to contemplate, at one point or another, some target proposition in probabilistic terms. This may happen in various different ways. Perhaps a more mathematics-adverse reasoner may sense simply his degree of belief in the truth or otherwise of such a proposition. But much theory and empirical evidence warns that such intuitive handling of probabilities (with no monitoring of adherence to formal rules) may exhibit demonstrably absurd behavior. Other reasoners may thus prefer a rigorous application of the rules of probability calculus including Bayes' theorem (in particular where the probability of a proposition in the light of new evidence needs to be evaluated). Usually, both kinds of reasoners will end up, at some point, with their personal states of belief about the truthstate of the source-level proposition at hand. Most likely, that belief will be a probability less than 1 and a variety of reasons can be forwarded to explain such a result. For example, the more 'intuitive reasoner' may hold that the source-level proposition relates to a past event of which the available evidence allows to give only an incomplete account (because the evidence may be degraded or there may be uncertainties in the analytical results). The more rigorously and probabilistically calculating reasoner, may add that his evaluative procedure will rely on probability estimates, such as a match probability or an error rate in the context of DNA evidence, that tend towards zero but never actually equal zero [13, e.g.].

Whichever position applies, a reasoner's probabilistic state of belief, as for itself, does not yet constitute a genuine individualization. Such a conclusion requires an additional step to be taken which consists of reconsidering the target proposition and assigning it a probability of 1 (certainty). This amounts to a deliberate suppression of uncertainty which, and most importantly, is *not* warranted by the evidence. This then is the point where the invoked arguments tend to become blurry. For example, it is sometimes said that the remaining uncertainty is so small that 'for practical applications it can be ignored'[6].

[6] Such statements still enjoy widespread popularity notably among parts of the firearms and tool mark identification community and are also inherent in the the-

In a Bayesian view, reaching a probability of unity would require an infinite likelihood ratio and the reasoner would also need to subscribe to statements of the kind 'however small the prior probabilities are, the likelihood ratio is so large that the posterior odds advance toward infinity' or 'the certainty will stand every amount of contrary evidence' [5, e.g.]. The point solely is that − assuming realistic circumstances − one cannot credibly arrive at a likelihood ratio that is infinite and from the ensuing statements just mentioned it is clear that these assume an extreme setting which it is doubtful to be true. In particular, one cannot conceive of any practical or theoretical means that would allow one to test, for example, whether every other person of the world is excluded as a potential source − apart from the general observation that such an open-set view is an excessively and, hence, often unnecessarily demanding assumption. Habitually, there is at least some information available that allows to narrow down the size of the population of potential sources [3], but this pertains to another discussion that is not further investigated here.

It thus appears that it is the very fact of deliberately suppressing uncertainty − which the established schemes need to achieve individualization − that breaks with a logical approach. A reasoner cannot, at one instance of time, rely on a probabilistic inferential procedure (however sophisticated and thoroughly rooted in rational considerations it may be) for evaluating the probability of a target source-level proposition and then, in just another instance of time, set the logical constraints of probability calculus out of control − in order to assign the target proposition a probability of one. This is radically contradictory. In addition, the suppression of uncertainty is − if it is done − ordinarily not accompanied by proper argument. This is why, in the view of Stoney [24], it is ultimately based on faith.

The tradeoff between an attempt to individualize and the lack of argument to support such a conclusion may be avoided, or at least be made explicit, if one's probability of a target proposition is not subsequently manipulated in a disproportionate way. Instead, that probability could be part of an argument that clarifies the structure and the ingredients of the problem as one of decision making. This is actually the central proposal of the present paper. Following the analyses set forth in Section 3, the reasoner faced with the question of individualizing or otherwise would have to make up his mind and valuate possible consequences of each decision. The final decision is then a function of relative desirability he assigns to the various consequences and his uncertainty about the true state of affairs (the suspect's (some other person's) being the source of the crime mark). Such an approach to forensic identification makes no reference to faith. Rather, it is an appeal to key players that engage in decision processes to assume their responsibility to explicate their probabilistic beliefs along with their preferential matrix.

---

ory of identification defined by the Association of Firearms and Toolmark Examiners (AFTE) [25].

5.2 Whose decision?

Throughout this paper, the discussion made reference to a hypothetical individual that is faced with a decision problem. It has intentionally been avoided to make an assumption of whether that individual is a scientist or any other actor involved in legal proceedings essentially because such an assumption is irrelevant for the argument's logical underpinnings.

But who, then, is to decide? The question of *who* ought to declare an identification appears to be as much debated as the question of *how* such a conclusion may be arrived at. According to experts in the field, there is reason to regard the 'traditional' approach (which rounds a near-one probability to one) to forensic identification both cautiously and skeptically. To quote Buckleton:

> My feeling is that we would be unwise to conclude the same source because it is not our place to do so. And if we do so, I would prefer the standard to be much higher than previously suggested AND I would like us to make transparent that we have *subjectively decided* to round a probability ESTIMATE off to zero. On balance I cannot see much positive coming from a policy of declaring a common source. [13, p. 109] [*italics added by the authors, capitals as in the original*]

This quote is interesting because it shows that interpreting an identification as a decision, independently of the individual who actually makes the decision, is already inherent in current discussions on the topic. There is yet another instance where the same author has crystallized this view as follows:

> To conclude the same source from a probabilistic model, *someone has to decide* that the probability estimate produced by that model at this extreme end of extrapolation is sufficiently reliable that it can be trusted and the probability is sufficiently small that it can be ignored. [13, p. 105] [*italics added by the authors*]

These quotes (and others of the same kind that may readily be found) also state that the decision involved in forensic identification is subjective, where 'subjective' is not meant to convey 'arbitrary' but 'personal' and 'related to a specific individual'.

The decision theoretic analyses proposed in Section 3 encapsulate these views with the sole but important difference that no debatable suppression of uncertainty is involved. In addition, and as a main contrast to the 'traditional' approach, they allow one to give some viable directions of answer in reply to the question of *who* ought to decide about an identification. If, ultimately, the recipient of expert evidence must find himself convinced of an identification, then it is *his* personal preferential setting that matters, along with *his* personal belief state about the truth or otherwise of the suspect's (some other person's) being the source of the crime mark. It does not seem sensible for a scientist to anticipate a preferential setting or even using one with default values. That would be tantamount to usurp upon the domain of the fact finder.

At this point, it is interesting to note that the proposed argument joins an analogous conclusion arrived at in the context of the discussion of whether scientists should address posterior probabilities of, for instance, a source-level proposition. There is argument in support of the position that forensic scientists cannot credibly offer such probabilities, on pain of ignoring the prior probabilities of the recipient of expert evidence [26,27, e.g.]. Similarly, it seems inconceivable for a scientist to assume a preferential setting that relates to a decision that is outside his area of competence. When a scientist attempts to define either prior probabilities or desirabilities for consequences of decisions he may deprive the recipient of expert evidence from engaging in his own thought process.

On balance, however, the individual who seeks to decide about an identification faces greater obligations than the individual who is solely concerned with evaluating a probability. The latter requires, among other ingredients, 'only' a prior probability whereas the former requires a probability for the target proposition plus a specification of desirabilities for outcomes. In both cases, the assessment of all the ingredients should receive careful attention and should be made in the light of all the relevant information of a case. If this cannot be guaranteed, then there is reason to doubt the appropriateness of the respective decision for the stated purpose.

5.3 The 'identification fallacy'

The point mentioned at the end of the preceding Section is of vital importance and it is advisable to be aware of the difference between a state of nature and a decision that might be taken with regard to that state of nature. Both are conceptually distinct entities and a decision d_i to believe in a given state of nature θ_j has no effect whatsoever on the actual truth or otherwise of that particular state of nature.

For the purpose of illustrating this point, let there be a reasoner who finds himself with a probability $Pr(\theta_1 \mid E) < 1$, that is, the probability of a designated individual (e.g., a suspect) to be the source of a crime mark given particular evidence E. Based on a decision d_1 to individualize that suspect, that is, a statement of a personal degree of conviction that θ_1 is true, it is fallacious to conclude that θ_1 is actually true. The reason for this is that the assumption of independence set forth in Section 3.2 implies that:

$$Pr(\theta_1 \mid E, d_1) = Pr(\theta_1 \mid E) \ .$$

Thus, with a probability $Pr(\theta_1 \mid E) < 1$, the decision d_1 still runs the risk to amount to a false identification with a probability $1 - Pr(\theta_1 \mid E)$.

The above mentioned independence assumption can be understood by considering that, ordinarily, θ relates to a real-world event that happened in the past. Evidence generated by that particular event that may be recovered, examined and used to inform one's beliefs about θ. But it does not appear feasible to conceive of an argument to demonstrate that a change in the state of mind (that is, a decision) of an individual, at a subsequent instant of time, could be of any relevance to the respective happening in the past.

5.4 Forensic identification: a really all that important issue?

In Section 5.2 it has been argued that in a decision theoretic conceptualization of forensic identification, a recipient of expert information should not expect forensic scientists to be competent in providing informed and recipient tailored opinions regarding individualization (unless the scientist is given *and* uses his client's preferential system).

Recommending a decision requires an effort to make up seriously one's mind on preferences although, finally, that decision cannot have, following the discussion in Section 5.3, an effect on a state of nature. Thus, is forensic identification, as for itself, an endeavor worth the effort?

There are, of course, obvious situations where a thorough study of individualization is an unavoidable requirement. Typical examples of this are scenarios involving anonymous human remains where there is a need of death certificates to be issued.

For a juror in a criminal trial of a common one-offender case, however, the relevance of forensic identification may be a much less vital issue. Actually, it may even be practically nil. If a juror's aim is to draw an inference to a crime-level proposition (e.g., 'the suspect (some other person) is the offender'), he may do so on the basis of his beliefs about a source-level proposition θ of the kind 'the suspect is (some other person) is the source of the crime stain'. The very point is that a juror *can* actually do this without the need to know whether θ is in fact true or not. He may solely need to account for evidential relevance [28] (that is, an assessment that he is proficient to provide given sufficient circumstantial information) and coherently informed beliefs about θ. The latter requirement can be appropriately complied with by considering a likelihood ratio that forensic scientists are competent to provide [9,10].

The implications of this are twofold. Firstly, in contexts where the aim is to draw probabilistic inferences about a crime-level proposition, debates about how one ought to decide about the truth or otherwise of a source level-proposition (that is, forensic identification) amount to attempts to providing answers which are, at least conceptually, needless.

Secondly, relevant directions of further research are those that aim at eliciting appropriate likelihood ratio formulae including means to assess the component parameters of such formulae. This is from where recent examples of this [29,30, e.g.] draw their legitimacy.

5.5 Relation with existing works

Although forensic identification − in its form as a statement of certainty − is not a necessary requirement for an inference to a crime-level proposition of the kind 'the suspect (some other person) is the offender' (Section 5.4), the decision theoretic apparatus outlined in this paper is not without relevance for discussions on that particular topic.

Instead of a decision at the source-level, it may well be that, for example, a decision may need to be made at the crime-level, that is, a propositional level that addresses the suspect's guilt and that pertains to the burden of persuasion. In this sort of context, decision theory − as it has been interpreted in the present paper − is a latently present subject [7,31–33]. In legal literature, the idea of studying adjudicative questions in a decision theoretic perspective can be traced back at least forty years [34], for instance.

6 Conclusions

Forensic identification is currently understood as a statement of certainty that designates an individual or object as the source of a particular item of evidence − to the exclusion of all other members of a pool of potential sources. Reliance on well developed probabilistic inference procedures is state of the art of addressing the uncertainty in that process, that is, the evaluation of target propositions (such as of 'the suspect (some other person) is the source of the crime stain') which are thought about in the light of so-called identification evidence.

Definite conclusions of identification cannot, however, be based on purely probabilistic schemes unless unrealistic assumptions are admitted [7]. As a consequence, forensic identification is not, at present, considered as the result of a scientific process [5, e.g.]. Actually, forensic identification requires a suppression of uncertainty that goes beyond that which is covered by actually available evidence.

In an attempt to both making that impasse explicit and to propose a way ahead to cope with, this paper advocates a conceptualization of forensic identification as one of decision making. The proposed approach demands its user only modest concessions and builds upon existing probabilistic inference procedures as an integral part. Essentially, the analyses presented in this paper

[7] For example, an infinite likelihood ratio as a result of considering the probability of the evidence, given that it originates from an alternative source, to be zero.

suggest themselves as a study of argumentative implications of a collection of basic concepts that a reasoner is willing to accept.

The central ideas articulate themselves through the following two main points:

- A sharp distinction can be made between what one ought to think about a proposition (such as, for example 'the suspect (some other person) is the source of the crime stain') and what one actually decides to render as a conclusion, that is, a statement that formally associates – or otherwise – a suspect (some other person) with the crime stain. The former is a problem that pertains to probabilistic reasoning whereas the latter is one that applies to decision making.
- Interpreting an identification as a decision does not make any reference to faith in that an actual suppression of uncertainty is safe – a step that would necessarily be needed when forensic identification is confined to a purely probabilistic setting. Instead, the proposed approach is an appeal to *(i)* a reasoned analysis of one's preferential standpoint, the available evidence including any relevant prior knowledge and *(ii)* the decision maker's duty to elicit these assessments in a serious and transparent way.

Any forensic identification procedure meets uncertainty as an unavoidable element that cannot be dissociated from the individual decision maker. As such, uncertainty precludes the possibility of giving any general solution by which accurate conclusions of identification and exclusion may be made. Although a decision theoretic approach to forensic identification involves no claim to dissolve the fundamental problem of incomplete information and imperfect evidence, it provides at least a rigorous framework to explicate the inherent 'risk' that one is prepared to accept if one is to engage in particular conclusions. It is hoped that the analyses proposed in this paper could be of assistance to both, those who wish to commit themselves to forensic identification and those who seek to examine such practice carefully (e.g., defense lawyers).

A Appendix

Fundamental axioms that should regulate personal preferences among alternative consequences in a decision problem are widely discussed in [16,35]. These axioms, also known as axioms of rational behavior, offer a valuable aid in the construction of the utility function since they imply the following result (see also [36]).

Consider any consequences c, c_1, c_2 from the sample of consequences, such that

$$c_2 \succ c \succ c_1,$$

that is c_2 is preferred to c (the symbol $\succ$ stands for *preference* among consequences), while c is preferred to c_1. There must be a gamble in which you get c_2 with probability γ and c_1 with probability $(1-\gamma)$ which is of the same value as c. In other words, there must exist a unique $0 < \gamma < 1$ such that

$$c \sim \gamma c_2 + (1-\gamma)c_1,$$

where the symbol $\sim$ represents *equivalence* among consequences.

When this latter relation is satisfied, the utility of any consequence c can be computed as:

$$U(c) = \gamma U(c_2) + (1-\gamma)U(c_1). \tag{A.1}$$

Let us go back to the decision table (Table 2) and see how this result can be implemented to build an utility function that encapsulates personal preferences of a rational decision maker. The ranking among consequences assumed in Section 3.3 can be formalized as follows:

$$C_{11} \sim C_{32} \succ C_{21} \sim C_{22} \succ C_{31} \succ C_{12}$$

To begin the construction of the utility function, we choose the best consequence (in this case the pair C_{11} and C_{32}) and the worst consequence (C_{12}), and let $U(C_{11}) = U(C_{32}) = 1$ and $U(C_{12}) = 0$. The utilities of the remaining consequences, C_{21}, C_{22} and C_{31}, will be established by comparison with, respectively, the best and the worst consequences, by assigning a value which seems reasonable with respect to the fixed utilities. Let us start by assigning an utility value to C_{21}, the 'neutral' consequence. The preference ranking outlined above states that

$$C_{11} \succ C_{21} \succ C_{12},$$

C_{11} is preferred to C_{21}, while C_{21} is preferred to C_{12}. The utility of C_{21} can be quantified with the help of the gambling scheme illustrated above (Equation A.1). There must exists a γ such that the consequence C_{21} is equivalent to a gamble where the best consequence is obtained with γ and the worst consequence is obtained with probability $(1-\gamma)$. The utility of C_{21} can then be computed as:

$$U(C_{21}) = \gamma U(C_{11}) + (1-\gamma)U(C_{12}) = \gamma.$$

Determining such an γ may be difficult: what would make one indifferent between an inconclusive decision and a situation in which a false identification incurs? One might agree that to be indifferent, γ should be rather high. If the possibility to end with a false identification is not negligible, then one might reasonably expect that any decision maker would prefer the 'neutral' consequence: the indifference relation would become one of preference. For the purpose of illustration, imagine that $\gamma = 0.999$ is felt to be correct. Then

$$U(C_{21}) = 0.999.$$

Notice that, given the equivalence relation among C_{21} and C_{22}, the utility of C_{22} is also set to 0.999.

Let us now consider consequence C_{31} (false exclusion). The preference ranking outlined above states that

$$C_{11} \succ C_{31} \succ C_{12}$$

The utility of C_{31} will be quantified in the same way as presented above, that is in comparison with $U(C_{11})$ and $U(C_{12})$:

$$U(C_{31}) = \gamma U(C_{11}) + (1-\gamma)U(C_{12}) = \gamma.$$

For a behavior to be rational, the value of γ must necessarily be lower than the previous one since the decision maker is facing, on the one side, the same gamble, while on the other side, a less preferred consequence ($C_{21} \succ C_{31}$). Let us say that $\gamma = 0.99$ is felt to be correct. Then

$$U(C_{31}) = 0.99.$$

Note that the order relation in the space of consequences is preserved. Nevertheless, one might object that there is no assurance that guarantees the coherence of the quantified utility values. Stated otherwise, does this utility function reflect personal preferences? This question can be examined by comparing different combinations of consequences:

$$C_{11} \succ C_{21} \succ C_{31} \quad ; \quad C_{21} \succ C_{31} \succ C_{12}$$

Consider the case on the left, for instance. According to the illustrated gambling scheme,

$$
\begin{aligned}
U(C_{21}) &= \gamma U(C_{11}) + (1-\gamma)U(C_{31}) \\
0.999 &= \gamma + (1-\gamma)0.99.
\end{aligned}
$$

When solving this equation one obtains $\gamma = 0.9$. Now, if one believes that this value is correct, in the sense that one is indifferent between a neutral consequence and a gamble where one might have a false exclusion with probability 0.1, then the utility function is coherent. Otherwise, one needs to go back and check previous assessments. Likewise,

$$
\begin{aligned}
U(C_{31}) &= \gamma U(C_{21}) + (1-\gamma)U(C_{12}) \\
0.99 &= \gamma 0.999.
\end{aligned}
$$

Solving this equation yields $\gamma = 0.99$. Now, if one believes that this value is correct, in the sense that one is indifferent between a false exclusion and a gamble where one might have a false identification with probability 0.01, then the utility function is coherent.

References

[1] P. L. Kirk, The ontogeny of criminalistics, Journal of Criminal Law, Criminology and Police Science 54 (1963) 235–238.

[2] C. R. Kingston, Application of probability theory in criminalistics, Journal of the American Statistical Association 60 (1965) 70–80.

[3] C. Champod, Identification/individualisation, in: J. H. Siegel, P. J. Saukko, G. C. Knupfer (Eds.), Encyclopedia of Forensic Science, Academic Press, San Diego, 2000, pp. 1077–1084.

[4] C. Champod, I. W. Evett, A probabilistic approach to fingerprint evidence, Journal of Forensic Identification 51 (2001) 101–122.

[5] I. W. Evett, B. S. Weir, Interpreting DNA Evidence, Sinauer Associates Inc., Sunderland, 1998.

[6] B. Robertson, G. A. Vignaux, Probability - the logic of the law, Oxford Journal of Legal Studies 13 (1993) 457–478.

[7] D. Lindley, Making Decisions, 2nd Edition, John Wiley & Sons, Chichester, 1985.

[8] D. V. Lindley, Probability, in: C. G. G. Aitken, D. A. Stoney (Eds.), The Use of Statistics in Forensic Science, Ellis Horwood, New York, 1991, pp. 27–50.

[9] C. G. G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd Edition, John Wiley & Sons, New York, 2004.

[10] B. Robertson, G. A. Vignaux, Interpreting Evidence. Evaluating Forensic Science in the Courtroom, John Wiley & Sons, Chichester, 1995.

[11] R. Cook, I. W. Evett, J. Jackson, P. J. Jones, J. A. Lambert, A hierarchy of propositions: Deciding which level to address in casework, Science & Justice 38 (1998) 231–240.

[12] D. V. Lindley, The philosophy of statistics, The Statistician 49 (2000) 293–337.

[13] J. S. Buckleton, C. M. Triggs, Relatedness and DNA: are we taking it seriously enough?, Forensic Science International 152 (2005) 115–119.

[14] J. S. Buckleton, C. M. Triggs, C. Champod, An extended likelihood ratio framework for interpreting evidence, Science & Justice 46 (2006) 69–78.

[15] C. Champod, C. Lennard, P. Margot, M. Stoilovic, Fingerprints and Other Ridge Skin Impressions, CRC Press, Boca Raton, 2004.

[16] J. O. Berger, Statistical Decision Theory and Bayesian Analysis, 2nd Edition, Springer, New York, 1985.

[17] F. Taroni, S. Bozza, C. G. G. Aitken, Decision analysis in forensic science, Journal of Forensic Sciences 50 (2005) 894–905.

[18] F. Taroni, C. G. G. Aitken, G. Garbolino, A. Biedermann, Bayesian networks and probabilistic inference in forensic science, John Wiley & Sons, Chichester, 2006.

[19] A. P. Dawid, J. Mortera, V. L. Pascali, D. van Boxel, Probabilistic expert systems for forensic inference from genetic markers, Scandinavian Journal of Statistics 29 (2002) 577–595.

[20] P. Garbolino, F. Taroni, Evaluation of scientific evidence using Bayesian networks, Forensic Science International 125 (2002) 149–155.

[21] F. Taroni, S. Bozza, A. Biedermann, Two items of evidence, no putative source: an inference problem in forensic intelligence, Journal of Forensic Sciences 51 (2006) 1350–1361.

[22] I. W. Evett, L. A. Foreman, G. Jackson, J. A. Lambert, DNA profiling: A discussion of issues relating to the reporting of very small match probabilities, The Criminal Law Review (2000) 341–355.

[23] I. Inman, N. Rudin, Principles and Practice of Criminalistics, The Profession of Forensic Science, CRC Press, Boca Raton, 2001.

[24] D. A. Stoney, What made us ever think we could individualize using statistics?, Journal of the Forensic Science Society 31 (1991) 197–199.

[25] R. G. Nichols, Defending the scientific foundations of the firearms and tool mark identification discipline: responding to recent challenges, Journal of Forensic Sciences 52 (2007) 586–594.

[26] F. Taroni, A. Biedermann, Inadequacies of posterior probabilities for the assessment of scientific evidence, Law, Probability and Risk 4 (2005) 89–114.

[27] A. Biedermann, F. Taroni, P. Garbolino, Equal prior probabilities: can one do any better?, Forensic Science International 172 (2007) 85–93.

[28] D. A. Stoney, Relaxation of the assumption of relevance and an application to one-trace and two-trace problems, Journal of the Forensic Science Society 34 (1994) 17–21.

[29] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, D. Meuwly, A. Bromage-Griffiths, Computation of likelihood ratios in fingerprint identification for configurations of three minutiæ, Journal of Forensic Sciences 51 (2006) 1–1.

[30] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, A. Bromage-Griffiths, Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiæ, Journal of Forensic Sciences 52 (2007) 1–1.

[31] R. O. Lempert, Modeling relevance, Michigan Law Review 75 (1977) 1021–1057.

[32] D. H. Kaye, What is Bayesianism?, in: P. Tillers, E. D. Green (Eds.), Probability and Inference in the Law of Evidence, The Uses and Limits of Bayesianism (Boston Studies in the Philosophy of Science), Springer, Dordrecht, 1988, pp. 1–19.

[33] D. H. Kaye, Clarifying the burden of persuasion: what Bayesian decision rules do and do not do, The International Journal of Evidence & Proof 3 (1999) 1–29.

[34] J. Kaplan, Decision theory and the factfinding process, Stanford Law Review 20 (1968) 1065–1092.

[35] M. H. De Groot, Optimal Statistical Decisions, Wiley Classics Library Edition Published 2004, John Wiley & Sons, Inc., Hoboken, New Jersey, 1970.

[36] C. Robert, The Bayesian Choice. A decision-theoretic motivation, Springer-Verlag, New York, 1994.

Figures 1 to 3

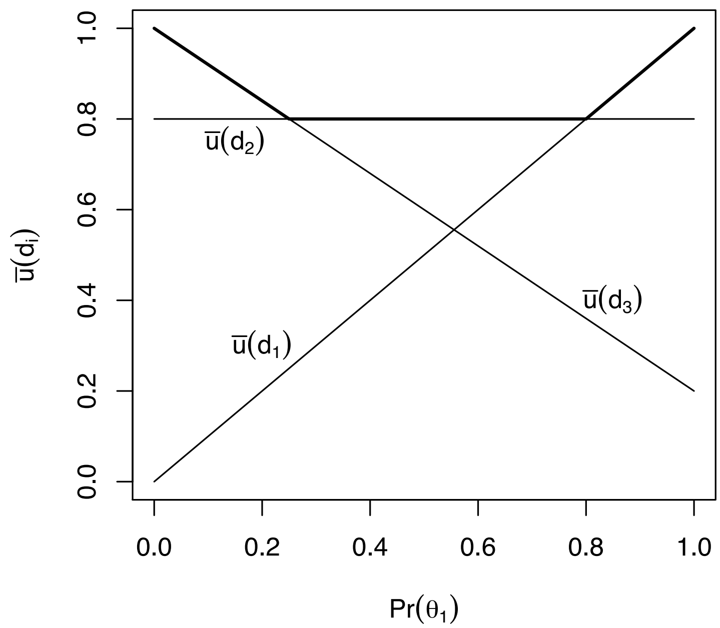


Fig. A.1. Expected utilities $\bar{u}(d_i)$ as a function of $Pr(\theta_1)$. Decision d_1 represents 'identification', d_2 'inconclusive' and d_3 'exclusion'. θ_1 is the proposition 'the suspect is the source of the crime mark'. The function plots represent Equations (3), (4) and (5) with α and β set to 0.8 and 0.2 respectively. The bold line marks the decision that maximizes expected utility given a particular (range of) value(s) for $Pr(\theta_1)$.

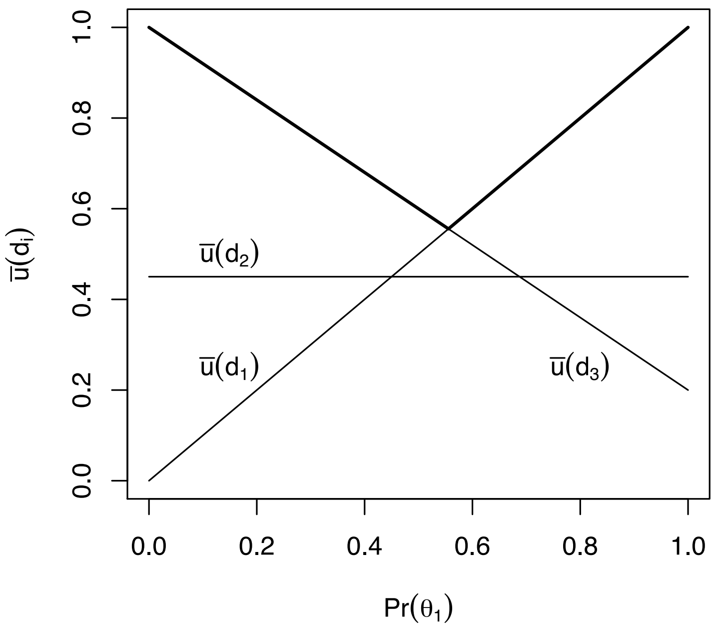


Fig. A.2. Function plots for Equations (3), (4) and (5) with α and β set to 0.45 and 0.2 respectively. The definition of the variables is the same as in Figure A.1. The bold line marks the decision that maximizes expected utility given a particular (range of) value(s) for $Pr(\theta_1)$.

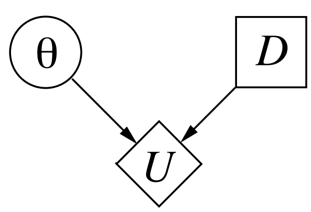


Fig. A.3. Influence diagram with a binary propositional variable θ (denoting 'the suspect (some other person) is the source of the crime mark'), a decision node D (covering d_1 for 'identification', d_2 for 'inconclusive' and d_3 for 'exclusion') and a utility node U.

Fingerprints at the crime-scene:

Statistically certain, or probable?

Fingerprints have been used for a century to identify criminals. But, astonishingly, fingerprint experts rely on subjective opinion, not on objective science. Yet they are required to claim absolute certainty for their judgements – a certainty that is mythical. **Cedric Neumann** brings probabilities and the hope of better justice to the courtroom; with Julian Champkin he explains the idea.

> With dramatic suddenness [Inspector Lestrade] struck a match and by its light exposed a stain of blood upon the whitewashed wall. … It was the well-marked print of a thumb. …
>
> "You are aware that no two thumb-marks are alike?"
>
> "I have heard something of the kind," replied Holmes.

Thus Sherlock Holmes in the case of 'The Norwood Builder', from *The Return of Sherlock Holmes*, first published in 1905. Since the dawn of modern detective work fingerprints have been regarded as the prime means of identifying an individual; their use as a means of personal identification goes back still further, to ancient China and even earlier (see box overleaf). Nowadays, fingerprints are used by police forces and courts worldwide. It is commonly believed, and may very well be true, that no two people share an identical fingerprint. But forensic scientists do not deal with exact reproduction, but have to draw conclusions from imperfect prints found at crime-scenes. The comparison of fingerprints is not an objective science. Fingerprint experts rely on judgements and opinion. They give their testimony in court; and the testimony they give is an entirely subjective one. Yet at the same time that testimony is required to be definitive and absolute.

The expert has a mark from a crime-scene, and a print taken from a known individual, and is required to say either that the mark matches the print, and was therefore made by the accused; or that it does not match and was made by someone else. Traditionally, no room for doubt is allowed: a fingerprint expert will report that he or she is *absolutely certain* that an impression from the scene of a crime comes from the finger of the accused.

But in real life nothing is certain. Fingerprint experts are exceedingly good at their jobs; nevertheless, different experts can and occasionally do give different opinions as to whether or not two impressions match. The crime-scene mark is unlikely to be perfect; it may be degraded, partial, distorted, or blurred. The expert may feel a small residue of doubt. If it is his belief that the mark 'probably does' or 'almost certainly does' or 'is rather unlikely to' match, he is forbidden to say so in court; in those cases, fingerprint evidence, for or against the accused, simply does not appear in the case and whatever supporting information content it may have is effectively wasted. (This happens in as many as 30% of the comparisons performed in a fingerprint bureau[1]). Strangely, DNA experts are *required* to give probabilities for their evidence of matching; fingerprint expert are *forbidden* to. This bizarre situation ought to be ended, in the interests of justice as well as of common sense.

DNA experts are *required* to give probabilities. Fingerprint experts are *forbidden*. This bizarre situation should be ended

The history of fingerprinting

By the year 246 BC Chinese officials were impressing their fingerprints into the clay seals used to seal documents. Some time before AD 851 an Arab merchant in China, Abu Zayd Hasan, witnessed Chinese merchants using fingerprints on silk or paper to authenticate loans. The Persian physician Rashid-al-Din Hamadani (1247–1318) refers to the Chinese practice of identifying people via their fingerprints and commented "Experience shows that no two individuals have fingers exactly alike."

Modern use of fingerprinting began with Sir William Herschel. In India, from 1858, he used fingerprints to authenticate legal documents. Around 1880, Dr Henry Faulds, working in Tokyo, proposed a fingerprint scheme to the Metropolitan Police in London, who rejected it. Faulds wrote to Charles Darwin. Darwin, being too old and ill to work on it, passed the information to his cousin, Sir Francis Galton, who devised the classification system of whorls, loops and arches still in use today. Galton's book *Fingerprints* (1892) encouraged its use in forensic science. He also performed one of the few statistical analyses of fingerprinting, calculating that the chance of two different individuals having the same fingerprints was about 1 in 64 billion. The first use of fingerprints to identify a murderer was in Argentina in 1892. Scotland Yard's Fingerprint Bureau was established in 1901.

If there is, say, a 75% probability of a match a jury should be told of that, to add or subtract from the balance of evidence against the accused, to weigh the scale on one side or the other. The present system, of a pretended certainty (one way or the other) that is not in fact there, can and should be improved upon.

What is needed is a way to give quantified, numerical figures to the probability of a match.

How, then, can we introduce probability to fingerprinting? The system itself has hardly changed since it was devised. The classifications that Francis Galton and Sir Edward Henry introduced, of loops, arches and whorls, are still used today. For a more objective and probability-based approach we need ways to quantify these swirling shapes.

First, some definitions: a *fingerprint*, or *print*, is the record that is taken, at the police station or somewhere similar, from a known individual, using ink pad and roller, under controlled conditions and protocols. It is, if you like, the gold standard: the impression is as clear as possible, and we know for certain who it comes from. A *mark*, on the other hand, is the impression that is found at the scene of the crime. It may be in blood, or grease, or powder; it may be smudged or smeared or distorted from a finger pressed at an angle; it may be incomplete; it may have other patterns – from wood grain, perhaps, if it is on a wooden surface – superimposed on it; it may be distorted from being on a curved surface such as a bottle or a glass. It will probably have been photographed, or lifted by powder and tape by a scene-of-crime expert using one of the techniques that crime-scene investigation shows portray rather fancifully on television. It will almost certainly not be as clear as our police-station fingerprint; and if you superimpose the two, almost certainly they will not exactly coincide.

So how can we establish whether police-station fingerprint and scene-of-crime mark comes from the same individual? And how can we give numerical probabilities for the certainty or uncertainty of our conclusion? The first problem is to find a set of features that define a fingerprint or mark. Then we can assign the probability that two similar but not identical sets of features come from the same source.

A fingerprint is essentially a pattern. It is too complex for mathematics or for fingerprint experts to analyse as a whole. Instead, experts select parts of it, which are deemed sufficient to distinguish it from any other fingerprint. Three basic overall patterns of fingerprint are arches, loops and whorls (see Figure 1), and these are good for general recognition and for classifying – as in the huge databases which police forces everywhere maintain; but for crime-scene comparisons the experts use finer details. Each of the few dozen ridges of skin that together form a fingerprint can have one of two things happen to it: it might come to an end; or it might divide into two (exceptionally three). These points of detail are called *minutiae*, or points, and it is by comparing the minutiae in a print and a trace that examiners form their opinions.

Quantitative observations of fingerprints

On the one hand, a single fingerprint pattern can contain dozens of these minutiae; there may be too many to examine them all. On the other hand, marks recovered from crime-scenes will only contain very few minutiae. When comparing marks and prints, experts will observe the minutiae present on the mark and search the print for correspondences and discordances. Historically, in most countries, 12 minutiae that matched each other in type, orientation and position have generally been considered sufficient to identify the source of the mark. Until 2001 the UK required 16 correspondences to establish proof of identity. Both these numbers arose through experience rather than statistical analysis.

Main patterns of ridge flow (◇, core (centre) of the patterns; △, delta (point of convergence of the ridge flow) of the patterns): (a) whorl; (b) arch; (c) loop

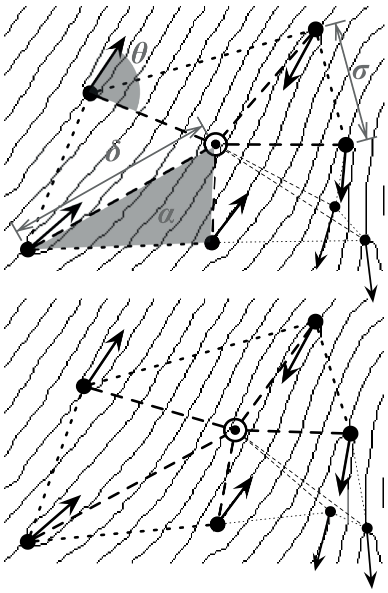


Figure 2. Illustrations of the radial triangulation used to order minutiae. The variables listed in Table 1 are indicated in grey in (a). The arrows show the 'directions' of the minutiae. The differences between (a) and (b) are presented by using thinner lines

Basil Rathbone as Sherlock Holmes, considering a probable match? Rathbone played Holmes in more than a dozen Hollywood movies from 1939 onwards

The reasoning that currently leads experts from minutiae to identification is essentially a psychological one that cannot be rationalized and rendered explicit. The method that my colleagues and I have presented[2] also relies on those minutiae; but numbers are derived from them.

On any given finger impression, the most prominent minutiae – say six – can be selected and joined up, in a clockwise direction (see Figure 2). They will form a pattern – essentially a six-sided polygon around a centre. (The centre can be defined as the arithmetic mean of the Cartesian co-ordinates of our six points.) A polygon is a much simpler pattern than the whirling lines of a full print or mark. It is also much easier to analyse numerically. The basis of the method is to describe that polygon with a set of variables.

If the polygon has k corners – that is, was constructed from k minutiae - it contains $5k$ pieces of numerical information. They are, for each corner: its distance from the centre; its distance from the next corner (going clockwise); the angle between the direction of the minutia and the radius line; the area of the triangle it forms with its neighbour and the centre; and the type of minutia it is – a ridge ending, a bifurcation, or unknown (Table 1). These numbers can be recorded as a column, and the columns for each minutia can be set side by side. So, if $k = 6$, the information on the fingerprint is reduced to a 6×5 array of numbers.

Let us return to the scene of the crime, and the perhaps partial and degraded mark we have found there. This too can be reduced to a 6×5 array of numbers. Those numbers may be similar to, but not identical to, the ones in the previous fingerprint array. When we compare the two there will be similarities and also differences. How likely is it that they come from the same finger? And how can we assign numerical probabilities to that vague term "likely"?

Comparison of configurations and optimization

The comparison of the two arrays can be visualized by returning to the polygons that generated them. If we lay the centres of the two polygons on top of one another we can rotate them until the distance between corresponding corners is minimized. A little maths and geometry will find the best position. We can add up all the six corner-differences to give a single number, or score, that expresses the similarities and differences between the two patterns. This is very similar in principle to matching algorithms implemented in the UK national fingerprint database known as Ident1.

The score is actually the differences between the numbers in the two arrays when they have been collectively minimized. The total difference between any two *minutiae* (corners) is the weighted arithmetic sum of the squared differences between the five variables measured on every minutia. The weightings are simple multipliers whose values are chosen to optimise the algorithm. The optimization criterion is to end up with a measure of distance such that mark–print pairs from the same finger (within source) should tend to yield small scores and mark–print pairs from different fingers (between source) should tend to yield large scores. Those interested may find the mathematics, the geometry and the weightings in the paper by Neumann et al. cited earlier[2].

Table 1. Data extracted from a minutiae configuation

Notation	*Description*	*Units*
δ	Radius – the distance between a *minutia* and the centre of the polygon	Pixels
σ	Side length – the distance between a *minutia* and the next contiguous *minutia* in a clockwise direction	Pixels
θ	Angle – the angle between the direction of a *minutia* and the line from the centre of the polygon to the *minutia*	Radians
α	Area – the area of the triangle constituted by a *minutia*, the next contiguous *minutia* and the centre of the polygon	Pixels
τ	Type – type of the *minutia* {ridge ending, bifurcation, unknown}	[0, 1, 2]

Properly interpreting the similarities and differences observed between a crime-scene mark and the control print from a suspect requires considering two alternative hypotheses. First, the hypothesis that the prosecution in a courtroom will try to prove: namely, that the trace and the source come from the same finger – that is, the differences can be explained by the partial and degraded nature of the mark. The method also needs to be fair to the defence, which would be trying to prove the opposite hypothesis, that the fingerprint and trace come from different fingers – that is, the similarities between the mark and the suspect's print are coincidental – and here we are interested in knowing how likely it is to randomly select an individual who will exhibit such similarities.

For use in courtrooms, we need a method to translate the scores calculated between marks and prints into probabilities assigned to the two hypotheses mentioned above. The first hypothesis assumes that the suspect is the real source of the crime-scene mark; it requires studying the range of possible marks that this suspect can leave, and assessing whether the mark recovered on the crime-scene fits within this range. Under this hypothesis, scores are calculated between the crime-scene mark and 'pseudo-marks' – that is, a large number of marks, generated by computer from the suspect's real fingerprint, each pseudo-mark varying from the real print according to algorithms that simulate possible distortions, pressures and angles of pressure and the like; the pseudo-marks represent the range of different traces that the suspect's actual finger could have made. They let us determine how likely it is to observe the crime-scene mark if it was truly left by the suspect. By definition of the score used in this study, a large amount of small scores will result in a high probability supporting this hypothesis.

The second hypothesis assumes that somebody other than the suspect is the real source of the crime-scene mark; in other words, it assumes that the suspect is wrongfully accused and that the similarities observed between the mark and his finger are coincidental. This hypothesis requires studying the number of people who have fingers that can generate a mark with the same features that are observed on the crime-scene mark, and assessing how rare or common are these features. Under this hypothesis, scores are calculated between the crime-scene mark and pseudo-marks generated from fingers of randomly selected individuals to determine how likely it is to observe fingers exhibiting the features of the crime-scene mark by random chance. By definition of the score used in this study, a large amount of small scores will result in a high probability of observing the features on the crime-scene mark by chance (common features) and support the hypothesis that the association between the crime-scene mark and the suspect is coincidental; while a negligible amount of small scores will result in a very low probability of observing those features by chance (rare features) and support the hypothesis that the association between the mark and the suspect is evidence in the case.

The ratio between the two probabilities is called a likelihood ratio. The higher the likelihood ratio, the stronger the evidence in favour of the hypothesis that the suspect *is* the source of the crime-scene mark; the smaller the likelihood ratio, the stronger the evidence in favour of the hypothesis that the suspect *is not* the source of the crime-scene mark.

Testing the model

The method needs to be tested and checked against reality. If print and trace come from the same fingers, the model must give a high probability supporting that hypothesis; if they come from different fingers, the model must reflect that fact equally strongly.

Ideally, we want neither that the method generates misleading evidence in favour of the prosecution, nor in favour of the defence. We want to be sure that if the numbers extracted from the mark and print polygons seem similar, this could not be accidental resemblance. To test this under the most difficult conditions, a validation experiment was performed: traces from the scenes of 364 crimes committed in North Wales were compared to fingerprints obtained from the national fingerprint database of the USA.

This US database contains over 600 million prints. Its experts searched it for the nearest match to each Welsh crime-scene trace; in each case, checks were made that the owner of the US finger concerned could not possibly be the source of the crime committed in North Wales. This gave us a gold standard of certainty with which to test our method's powers of discrimination.

In each case it was found that the US fingerprint which corresponded most closely – out of over 600 million of them – to the trace at the Welsh crime-scene was numerically different enough not to be giving misleading evidence that the US person was the source of the mark. In other words, a prosecution based on that fingerprint evidence would not have been brought based on the model. This was reassuring: it was a tough test, yet the method would in no case have helped to convict an innocent person. What of the reverse hypothesis? Fingerprint evidence is more usually used in court to help identify the guilty.

The North Wales crime-scene database was from cases that had been brought to court and successfully prosecuted; so the fingerprints from the true British-based perpetrators of the crimes were also available to compare to the traces. Here, too, the model performed satisfactorily. There is a clear separation between the likelihood ratios in the cases where print and mark came from the same person, and the cases where they came from different people. The more minutiae were used, the clearer was the separation.

The future

In the immediate future, we do not see that the current courtroom practice of presenting categorical opinions about fingerprints will change. But in the longer term we expect an evolution towards a framework that is similar to that which underpins DNA evidence. Already this work has formed the basis of training workshops in the UK, USA and in Europe. We have seen several years of courtroom battles in relation to DNA evidence. They have proved to be beneficial to the science. We must expect similar battles over this method. But the notion of a quantitative measure of the strength or weakness of evidence involves subtle issues which many lawyers fail to understand. It remains to be seen how future legal battles play out, but we see models such as this one as a powerful platform for change.

References

1. Neumann, C., Mateos-Garcia, I., Langenburg, G., Kostroski, J., Skerrett, J.E. and Koolen, M. (2011) Operational benefits and challenges of the use of fingerprint statistical models: a field study. *Forensic Science International*, **212**, 32–46.
2. Neumann, C., Evett, I.W. and Skerrett, J. (2012) Quantifying the weight of evidence assigned to a forensic fingerprint comparison: a new paradigm. *Journal of the Royal Statistical Society*, *Series A*, **175**, 1–26.

Cedric Neumann is currently an Assistant Professor in Statistics at Pennsylvania State University. Cedric obtained his Ph.D. in Forensic Science from the oldest forensic programme in the world at the University of Lausanne, Switzerland. He has worked on forensic projects for the United States Secret Service and has managed a research team at the Forensic Science Service in the UK.

Scientific Working Group on Friction Ridge Analysis Study and Technology, "Document # 10: Standards for Examining Friction Ridge Impressions and Resulting Conclusions" (2011) 1-16. Document in public domain.

SWGFAST

Scientific Working Group on Friction Ridge Analysis, Study and Technology

Document #10

Standards for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint)

1. Preamble

1.1. Friction ridge impression examinations are conducted by examiners using the Analysis, Comparison, Evaluation, and Verification (ACE-V) methodology, which include both qualitative and quantitative aspects. ACE is not generally applied as a strictly linear process because it may include a return to any previous phase. Application of ACE includes observations, measurements, assessments, decision-making, and documentation, which are enabled by the education, training, skill, and experience of the examiner.

1.2. The examination of friction ridge impressions and the resulting conclusions are based on ridge flow and ridge paths; the location, direction, and spatial relationships of minutiae; and ridge structure. The analysis phase leads to the determination of *suitability.* Following comparison, the evaluation phase leads to the following conclusions: *individualization*, *exclusion*, or *inconclusive*. These conclusions are based on the following premises [1] [2]:

1.2.1. Friction ridge skin bears an extremely complex, unique, and persistent morphological structure.

1.2.2. Notwithstanding the pliability of friction ridge skin, the contingencies of touching a surface, and the nature of the matrix, an impression of friction ridge skin structure may be left following contact with a surface.

1.2.3. This impression may display features of varying quality (clarity of ridge features) and specificity (weighted values and rarity).

1.2.4. Notwithstanding variations in clarity and specificity, the unique aspects of friction ridge skin may be represented as highly discriminative features in impressions.

1.2.5. An impression that contains sufficient quality and quantity of friction ridge features can be individualized to, or excluded from, a source.

1.2.6. The use of a fixed number of friction ridge features as a threshold for the establishment of an individualization is not scientifically supported.

2. Scope

2.1. The ACE-V methodology of friction ridge impression examination utilizes a qualitative and quantitative assessment of Level 1, Level 2, and Level 3 details.

2.2. The ACE-V methodology is applied to examinations and comparisons of friction ridge impressions. This document illustrates the case of unknown to known comparisons but is applicable to other comparisons (e.g., known to known).

2.3. The application of the ACE-V methodology to casework requires examiner competency as established through training and testing [3].

3. Factors Affecting Examinations

3.1. The following factors affect the qualitative and quantitative aspects of friction ridge impressions. A competent examiner [3] will understand these factors, recognize that they occur in friction ridge impressions, and understand how they influence friction ridge impression reproducibility. These factors may cause an apparent dissimilarity between impressions from the same source. Failure to properly assess the occurrence and influence of these factors could result in misinterpretation. When applicable, the following factors must be considered in all steps of the ACE-V methodology:

3.1.1. Anatomical aspects include the condition of the skin (e.g., scars and warts) and the morphology of the hand and foot relative to the shape and contour of the substrate.

3.1.2. Transfer conditions include pressure applied during transfer, slippage or twisting, sequence of deposition (i.e., double taps and overlays), and an understanding of the limitations of friction ridge pliability.

3.1.3. Matrix includes bodily secretions and contaminants (e.g., sweat, blood, paint, dirt, oil, grease).

3.1.4. Detection techniques that can be one or more of the following: optical (i.e., light sources and illumination techniques), physical, or chemical processing techniques.

3.1.5. Recording or preservation techniques, such as photography, lifting, live-scan, and ink.

3.1.6. Substrate (e.g., porous, non-porous, semi-porous, smooth, rough, corrugated, pliable, or textured surfaces).

3.1.7. Environmental conditions (e.g., protected, unprotected, wet, dry, cold, or hot).

4. Levels of Friction Ridge Impression Detail For Examinations

Level 1 detail refers to the overall ridge flow. Level 2 detail refers to individual friction ridge paths, friction ridge events (e.g., bifurcations, ending ridges, dots, and continuous ridges), and their relative arrangements. Level 3 detail refers to ridge structures (edge shapes and pores), and their relative arrangements. Creases, scars, warts, incipient ridges, and other features may be reflected in all three levels of details[1].

5. Procedure for Friction Ridge Impression Examinations (ACE-V Methodology)

5.1. Analysis

5.1.1. Analysis includes the assessment of the impression to determine its value based on Level 1, 2, and 3 detail. This assessment is affected by other relevant information as described in section 2, as well as possible anatomical origin and orientation. Analysis determines if the impression is suitable for comparison. If the impression is not suitable, the examination will stop at the analysis phase and will be reported as such[2]. If the impression is suitable, the analysis further indicates the features and their tolerances to be used in the comparison.

5.1.2. In the analysis phase, the examiner assesses the friction ridge skin features and determines the tolerances assigned to the impressions (unknown and known). Tolerance is the allowance of variation in appearance of friction ridge features (due to the factors listed in section 3) that will be accepted during comparison, should the corresponding print be available.

[1] For example, a crease could exhibit Level 1 crease flow, Level 2 crease path, and Level 3 crease shape.

[2] It will not be reported as inconclusive, but may be submitted to verification.

5.1.3. The analysis may also provide anatomical information to prioritize the potential corresponding areas and limit unnecessary comparisons. Certain orientation indicators such as recurves, deltas, creases, and scars may provide specific guidance where to begin the comparison.

5.1.4. Determination of suitability

5.1.4.1. The determination of suitability is based on the assessment of the discriminating strengths of the features and their arrangements. Suitability is the determination that there is adequate quality and quantity of friction ridge features in an impression for some further process step. The assessment is made based on the quality of features (clarity of the observed features), the quantity of features (amount of features and area), the specificity of features, and their relationships (see section 5).

5.1.4.2. There are commonly two approaches to the determination of suitability often adopted as agency policy:

5.1.4.2.1. Approach #1 (commonly referred to as “of value for identification”): Only impressions of *value for individualization* are compared. *Value for individualization* indicates an impression that is deemed to be identifiable. When adopting this approach, impressions lacking value for individualization are not further compared.

5.1.4.2.2. Approach #2 (commonly referred to as “of value for comparison”): Impressions of value for individualization and impressions only of value for exclusion are compared.

5.1.4.2.3. Conclusions in the evaluation phase following both approaches are: individualization, exclusion, or inconclusive.

5.2. Comparison

5.2.1. If the analysis phase provides indicators as to the probable anatomical area, a side-by-side comparison with the appropriate area of the known print is initially conducted. In the absence of indicators, all areas of available known impressions must be compared.

5.2.2. Comparison is accomplished through the side-by-side observation of all levels of details to determine whether the two impressions are in agreement or disagreement based upon features, sequences, and spatial relationships within the tolerances of clarity and distortion.

5.2.3. Comparison begins with the determination of dissimilarity or similarity between two impressions at Level 1. If similarity is determined within tolerance at Level 1, a target group is selected from the features observed during the analysis phase and is then searched within the selected area of the other impression. When similarity with the target group exists, additional contiguous arrangements of features are compared between impressions in a cyclical or recurring process from the unknown to the known impression to evaluate disagreement or agreement between the impressions. The process can be extended to comparing features in the known with features in the unknown that were reanalyzed during the comparison phase. If the initial target group is not found, alternative target groups may be selected and compared.

5.2.4. Observation of agreement or disagreement between the impressions initiates the evaluation phase.

5.3. Evaluation

5.3.1. Once the examination progresses from the comparison phase into the evaluation phase, it is determined whether the information is sufficient (see section 6) to form one of the three conclusions or return to the analysis phase and reassess suitability[3].

[3] This would not be necessary under approach 2.

5.3.2. In the evaluation phase, the examiner will ultimately decide whether the unknown impression is from a different source or the same source as the compared impression, or is inconclusive. These conclusions are defined below.

5.3.2.1. Exclusion

Exclusion is the decision by an examiner that there are sufficient features in disagreement to conclude that two areas of friction ridge impressions did not originate from the same source. Source refers to the area of friction skin. Exclusion of a subject can only be reached if all relevant comparable anatomical areas are represented and legible in the known exemplars. Notes and reports shall clearly state if the exclusion refers only to the source or the subject.

5.3.2.2. Individualization

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

5.3.2.3. Inconclusive

5.3.2.3.1. An inconclusive conclusion resulting from a suitability decision as described in approach #1 in section 5.1.4.2.1 occurs when an examiner is unable to individualize or exclude due to an absence of complete and legible known prints (e.g., poor quality fingerprints and lack of comparable areas). In such an instance, the inconclusive conclusion means that the impression needs to be reexamined using clearly and completely recorded known impressions.

5.3.2.3.2. An inconclusive conclusion resulting from a suitability decision as described in approach #2 in section 5.1.4.2.1 can occur either as in approach #1 or when corresponding features are observed but not sufficient to individualize. Likewise dissimilar features may be observed but not sufficient to exclude. In either case, the inconclusive conclusion means that the unknown impression was neither individualized nor excluded as originating from the same source.

5.3.2.3.3. There may be other instances where agencies have adopted procedures to report inconclusive conclusions. These are left to the administrative policies and procedures of the individual agency. However, these policies and reporting procedures must be clearly defined by the agency.

5.3.3. Reporting Conclusions

The conclusions of individualization and exclusion will be documented in notes and in reports; however, the determining factors need not be included in reports. Reasons for reaching inconclusive conclusions must be documented in notes and included in reports.

5.4. Verification

5.4.1. The independent application of the ACE process is utilized by a subsequent examiner to either support or refute the conclusions of the original examiner.

5.4.2. Suitability determinations may be verified by another examiner trained to competency [3]. A conclusion of individualization shall be verified. All other conclusions resulting from the evaluation phase should be verified.

5.4.3. Conflict resolution shall take place if the original conclusion is contested and cannot be resolved through consultation [5].

5.5. The flowchart in Appendix A details the major steps of ACE-V. The chart has been adapted from the NIST (National Institute of Standards and Technology) Expert Working Group on Human Factors in Latent Print Analysis. It is offered here as supporting documentation and applies to both tenprint and latent print examination.

6. Sufficiency for Conclusions

6.1. Sufficiency is a product of the quality and quantity of the objective data under observation (e.g., friction ridge, crease, and scar features). As the quality of an impression increases the need for quantity of friction ridge features decreases, as well as the inverse.

6.2. Quality

6.2.1. Quality is the assessment of the clarity of ridge features. Generally as quality increases so does the discernability and reliability[4] of the ridge features. It is recognized that quality is not necessarily constant throughout an impression. The assessment of quality may represent just the areas of highest quality, a range of qualities, or a map or rating system of quality of various regions in a single impression.

6.2.2. Table 1 shall be used for categorizing the levels of quality of the features in an impression (unknown or known). The level of quality determines the degree of tolerances that will be used during the comparison process. High quality will lead to low tolerances and conversely low quality will require high tolerances[5].

Quality	
High	Level 1 is distinct; Level 2 details are distinct; There are abundant distinct Level 3 details.
Medium High	Level 1 is distinct; Most of the Level 2 details are distinct; There are minimal distinct Level 3 details.
Medium Low	Level 1 is distinct; Few of the Level 2 details are distinct; There are minimal distinct Level 3 details.
Low	Level 1 may not be distinct; Most of the Level 2 details are indistinct; There are no distinct Level 3 details.

Table 1: Categories of quality defined as a function of levels of details observed.

[4] Reliability refers to the confidence assigned by the examiner to the observed ridge features in terms of existence, location, and shape he or she would expect to be reproduced on the corresponding print, should it be made available.

[5] High tolerances: generous allowances for variations in appearance and spatial relationships.

6.2.3. The above quality metric was designed to allow for a range of quality assessment as opposed to a narrow categorization. Table 1 provides four ranked categories for the quality metric. There are subjective as well as objective elements to this categorization, but the descriptions provided in the table should allow a meaningful quality description to be made with reference to the categories.

6.3. Quantity

6.3.1. Quantity, as applied in this section, is the number of ridge endings, bifurcations, and dots (minutiae) in contiguous ridges, determined without any reference to known impressions. All minutiae are considered here including indistinct minutiae for which type or exact location cannot be established[6]. Overall quantity of all features in the impression is not part of this measure.

6.3.2. It is recognized that this is an incomplete measure of the overall quantity of detail in a print. Level 2 detail encompasses more than minutia counts (including the ridge path, areas with open fields, and selectivity of minutiae). Minutia counts remain, however, as a discrete, measurable aspect of all prints and their enumeration is part of the systematic, formal consideration of quantity.

6.3.3. The utility of the number of minutiae as applied in this section is to assist in the analysis of suitability and the recognition of alternative levels of case complexity as they relate to sufficiency, evaluation, and verification. This use of the number of minutiae should not be considered as suggesting or endorsing the use of minutia counts as the sole criteria for a decision threshold.

6.4. Decision-Making

6.4.1. Sufficiency graph

6.4.1.1. The sufficiency graph (Figure 1) reflects the interplay between quality (defined in Table 1) and quantity of minutiae (as discussed in section 6.3) and its relation to the decision thresholds and levels of complexity based on a consensus of collective experience. It broadly represents how the amount of available information in an impression directly impacts the decision-making process. The sufficiency graph was developed to illustrate the intellectual process involved with the examination of friction ridge detail and the ensuing decisions. It represents the examiner's understanding of the aggregate relationship of details. Its purpose is to illustrate a part of the process dealing with the analysis of the impression for sufficient quality and quantity of detail to proceed with the comparison effort. It also illustrates certain thresholds wherein examiners should recognize the need for, and provide, enhanced documentation supporting their conclusions.

6.4.1.2. The axes used to plot the decision of the examiner, the positions of the curves, and the underlying regions, were created based on a consensus of experienced examiners (SWGFAST). Considerations in establishing the graph are related to actual casework and include international practices, general awareness of longstanding, as well as current literature and trends in ongoing research.

6.4.1.3. Level 2 detail in this graph is represented on the horizontal axis by numbers of minutiae. The limitations and rationale for using this metric for quantity are discussed in section 6.3. It is re-emphasized here that this should not be considered as suggesting or endorsing the use of minutiae counts as the sole criteria for a decision threshold.

---

[6] For example when a single ridge flows into a visually obscured area and two ridges emerge from the same area.

6.4.1.4. The four categories of quality represented on the vertical axis are given in Table 1 and discussed in section 6.2.

6.4.1.5. In Figure 1, the solid curve in the graph defines the lower limit of the sufficiency of friction ridge details below which, in area marked **A**, an individualization decision is not warranted. The dotted curve indicates the boundary between levels of complexity (complex versus non-complex). In area marked **B** in Figure 1, the examination is considered as complex and an individualization may be warranted. In area marked **C** in Figure 1, the examination is considered as non-complex and an individualization is warranted.

6.4.1.6. Quantity is meaningless in the absence of quality. Individualization cannot be achieved on quantitative considerations alone. It is recognized that in the absence of any minutiae, an individualization may be possible in such complex cases if the impression displays very high quality.

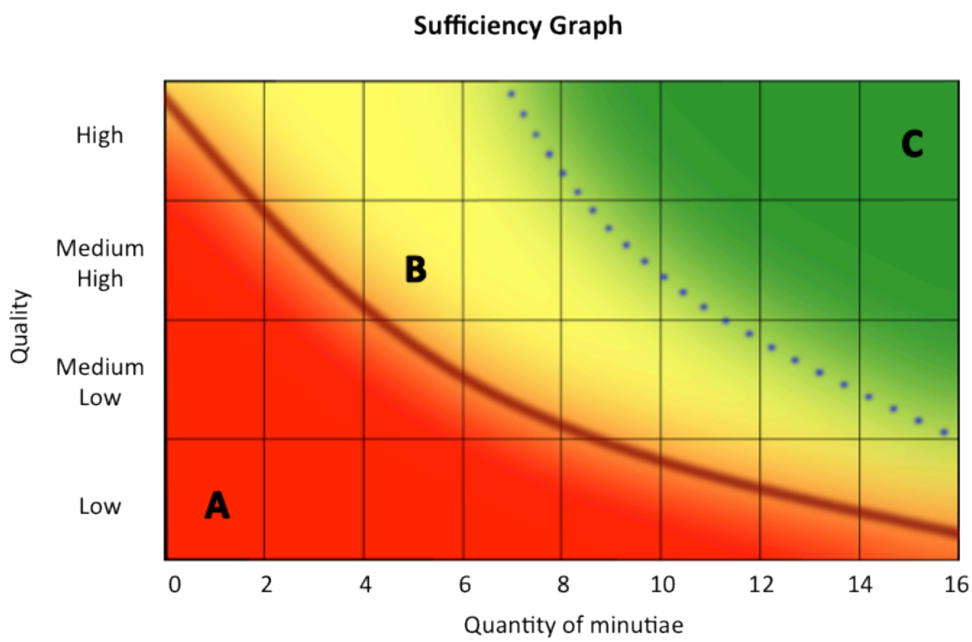


Figure 1: Sufficiency Graph

This graph does not suggest or endorse the use of minutiae counts as the sole criteria for a decision threshold.

6.4.2. Analysis phase

6.4.2.1. In the analysis phase, the assessment of the impression based on quality and quantity (as defined above) is positioned on the graph to determine its suitability for individualization. If the impression falls below the solid curve, then an individualization is not warranted. If positioned above the curve, then it may allow an individualization.

6.4.2.2. Minimum quality assurance measures are associated with each level of complexity according to the following table (Table 2):

Documentation [4] and verification procedures [5]	
Non-complex	Limited documentation of the relevant features used as a basis for a conclusion. Standard verification.
Complex	Extensive documentation of the relevant features used as a basis for a conclusion. Should consider the possibility of an enhanced verification and review procedure (e.g., a blind verification and multiple verifiers).

Table 2: Documentation and verification procedures.

6.4.2.3. A non-complex impression may be classified as complex if the following modifying factors are present: low specificity of features, significant distortion (e.g., multiple tap, superimposed impression, extreme pressure leading to tonal reversal, and slippage), high tolerances, or the original conclusion is contested during verification.

6.4.2.4. An impression categorized initially as complex may be classified as non-complex if modifying factors are present such as high specificity of features, presence of creases, scars, and open fields.

6.4.2.5. Justification for reassignment of complexity shall be documented.

6.4.3. Evaluation phase

6.4.3.1. In the evaluation phase, the sufficiency graph is used as a guide that broadly delineates the boundaries between individualization and inconclusive decisions.

6.4.3.2. In the evaluation phase, the decision process starts with an attempt at exclusion followed by an assessment of the potential correspondence observed between the impressions.

6.4.3.2.1. Exclusion (See Appendix B)

6.4.3.2.1.1. An exclusion decision can be based solely on Level 1 when sufficient pattern area and orientation indicators (e.g., recurves, cores, deltas, and creases) are available and when disagreement has been observed absent any significant distortion such as: double tap, overlaid impressions, or twisting. If significant distortion is observed, an exclusion decision can only be reached by considering both Level 1 and Level 2 details. If available, Level 3 detail may also be considered in conjunction with Level 2 detail.

6.4.3.2.1.2. An exclusion decision can be based on Level 2 detail when sufficient disagreement has been observed.

6.4.3.2.1.3. Level 3 details cannot be the sole factor in exclusion decisions. Level 3 details have to be considered in conjunction with Level 1 and Level 2 details.

6.4.3.2.2. Individualization

6.4.3.2.2.1. If the impressions originated from the same source, the examiner should observe correspondence, within tolerances, of all features. It is the degree of correspondence of features that is weighed.

6.4.3.2.2.2. For an individualization conclusion, sufficient agreement of information must exist so that the likelihood the impression was made by a different source is so remote that it is considered as a practical impossibility.

6.4.3.2.2.3. Level 3 details cannot be the sole factor in individualization decisions. Level 3 details have to be considered in conjunction with Level 1 and Level 2 details.

6.4.3.2.3. If the examiner did not reach a conclusion of individualization or exclusion, the conclusion will be reported as inconclusive.

7. References

[1] IAI 2010-18 resolution.

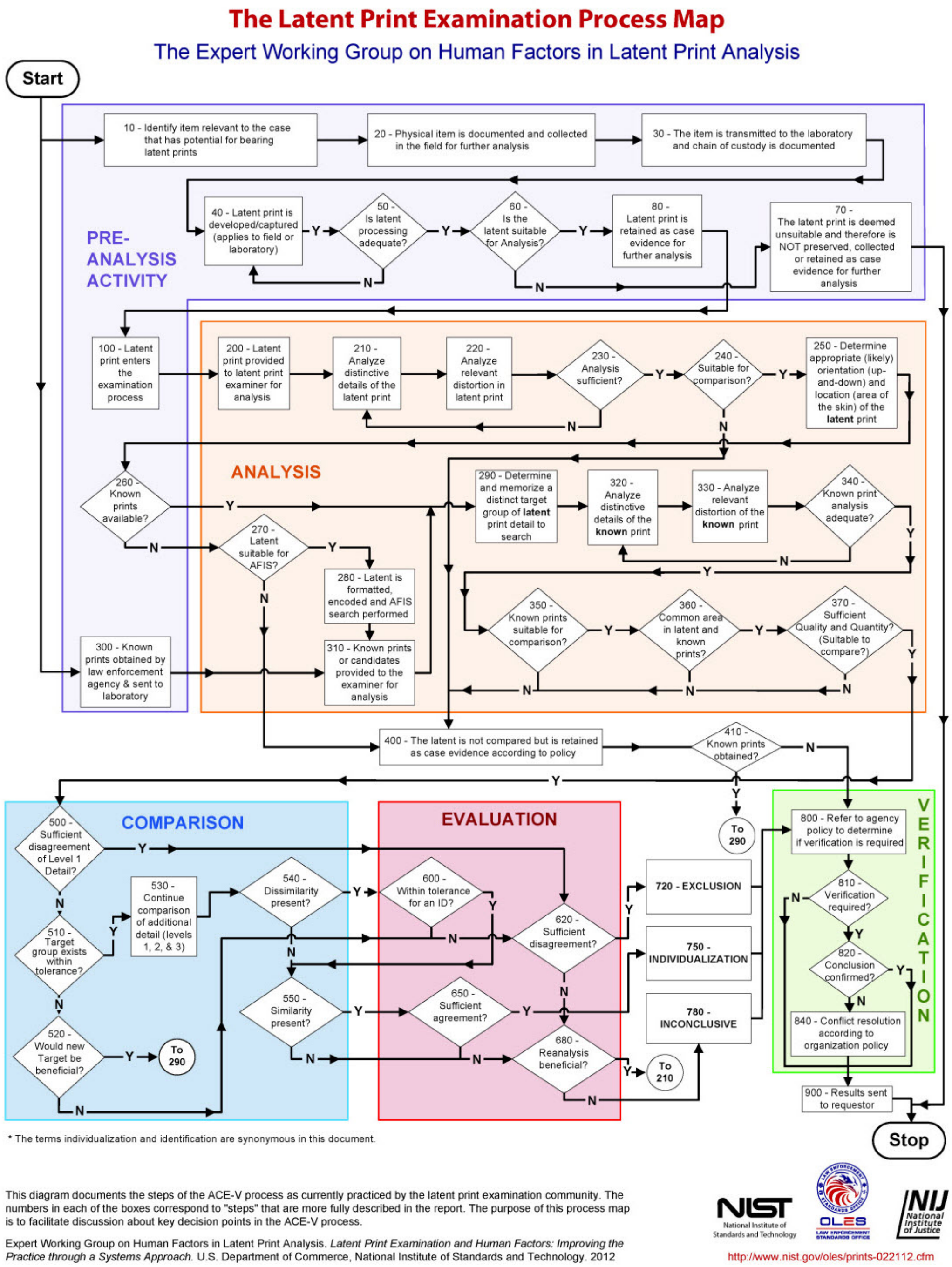
[2] Polski, J.; Smith, R.; Garrett, R. The Report of the International Association for Identification, Standardization II Committee. Grant no. 2006-DN-BX-K249 awarded by the U.S. Department of Justice, Washington, DC, March 2011.

[3] SWGFAST Document #18 (2010), Standards for Minimum Qualifications and Training to Competency for Friction Ridge Examiner Trainees (Latent/Tenprint), Ver. 2.0.

[4] SWGFAST Document #8 (2012), Standard for the Documentation of Analysis, Comparison, Evaluation, and Verification (ACE-V) (Latent), Ver. 2.0.

[5] SWGFAST Document #7 (2012), Standard for a Quality Assurance Program in Friction Ridge Examinations (Latent/Tenprint), Ver. 5.0.

Appendix A



The Latent Print Examination Process Map
The Expert Working Group on Human Factors in Latent Print Analysis
Start
PRE-ANALYSIS ACTIVITY
10 - Identify item relevant to the case that has potential for bearing latent prints
20 - Physical item is documented and collected in the field for further analysis
30 - The item is transmitted to the laboratory and chain of custody is documented
40 - Latent print is developed/captured (applies to field or laboratory)
50 - Is latent processing adequate?
60 - Is the latent suitable for Analysis?
80 - Latent print is retained as case evidence for further analysis
70 - The latent print is deemed unsuitable and therefore is NOT preserved, collected or retained as case evidence for further analysis
100 - Latent print enters the examination process
ANALYSIS
200 - Latent print provided to latent print examiner for analysis
210 - Analyze distinctive details of the latent print
220 - Analyze relevant distortion in latent print
230 - Analysis sufficient?
240 - Suitable for comparison?
250 - Determine appropriate (likely) orientation (up-and-down) and location (area of the skin) of the latent print
260 - Known prints available?
270 - Latent suitable for AFIS?
280 - Latent is formatted, encoded and AFIS search performed
290 - Determine and memorize a distinct target group of latent print detail to search
320 - Analyze distinctive details of the known print
330 - Analyze relevant distortion of the known print
340 - Known print analysis adequate?
300 - Known prints obtained by law enforcement agency & sent to laboratory
310 - Known prints or candidates provided to the examiner for analysis
350 - Known prints suitable for comparison?
360 - Common area in latent and known prints?
370 - Sufficient Quality and Quantity? (Suitable to compare?)
400 - The latent is not compared but is retained as case evidence according to policy
410 - Known prints obtained?
To 290
COMPARISON
500 - Sufficient disagreement of Level 1 Detail?
510 - Target group exists within tolerance?
520 - Would new Target be beneficial?
530 - Continue comparison of additional detail (levels 1, 2, & 3)
540 - Dissimilarity present?
550 - Similarity present?
To 290
EVALUATION
600 - Within tolerance for an ID?
620 - Sufficient disagreement?
650 - Sufficient agreement?
680 - Reanalysis beneficial?
To 210
720 - EXCLUSION
750 - INDIVIDUALIZATION
780 - INCONCLUSIVE
VERIFICATION
800 - Refer to agency policy to determine if verification is required
810 - Verification required?
820 - Conclusion confirmed?
840 - Conflict resolution according to organization policy
900 - Results sent to requestor
Stop
Y
N
* The terms individualization and identification are synonymous in this document.
This diagram documents the steps of the ACE-V process as currently practiced by the latent print examination community. The numbers in each of the boxes correspond to "steps" that are more fully described in the report. The purpose of this process map is to facilitate discussion about key decision points in the ACE-V process.
Expert Working Group on Human Factors in Latent Print Analysis. Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach. U.S. Department of Commerce, National Institute of Standards and Technology. 2012
NIST National Institute of Standards and Technology
OLES Law Enforcement Standards Office
NIJ National Institute of Justice
http://www.nist.gov/oles/prints-022112.cfm

Appendix B

8. **Exclusion of Source or Subject**

8.1. Exclusion is the decision by an examiner that there are sufficient features in disagreement to conclude that two friction ridge impressions originated from different sources. Exclusion implies that the likelihood of observing these features in disagreement, if the impressions are coming from the same source, is so remote that it is considered as a practical impossibility.

8.2. Excluding a known friction ridge impression as the source of an unknown impression requires that the examiner has determined that the anatomical region indicated in the unknown impression is fully represented in the exemplar. Exclusion of a subject as the origin of the unknown impression can only be reached if all relevant comparable anatomical areas are represented and legible in the known exemplars. Notes and reports shall clearly state if the exclusion refers only to the source or the subject.

8.3. When the impressions are determined to be non-complex, as determined in the analysis phase, exclusion can be simple. For example when the fingerprint is an impression of an arch pattern and the specific known impression under consideration is a whorl pattern the exclusion of that source is warranted (Figure B1).

Figure B1: High quality impressions with unambiguous features may be excluded as having a common source based on Level 1 detail.

8.4. The "One Discrepancy Rule" or the "One Dissimilarity Doctrine" as described by Thornton [Appendix B: 1], has been used to support exclusions of a source to an unknown impression. Although sometimes used synonymously and applied inconsistently, the terms "discrepancy" and "dissimilarity" refer to different concepts. Discrepancy refers to the presence of one or more friction ridge details in one impression that do not exist in the corresponding area of another impression [Appendix B: 2]. Discrepancies originate in the source skin. The term discrepancy is only used as a description of incompatibility between two impressions that has resulted in a conclusion of exclusion. It has long been recognized by latent print examiners that this "rule" should be critically applied with a detailed analysis of the remainder of the impression. Dissimilarity refers to a difference in appearance between two friction ridge impressions. For example, in the presence of overwhelming correspondence of features supporting the conclusion of individualization, an isolated dissimilarity may not be sufficient to exclude. Such a dissimilarity may be accepted as an artifact of distortion in the print or scarring in the skin without the examiner knowing the actual cause of the dissimilarity.

8.5. The "One Discrepancy Rule" has been historically associated with minutiae and was applied only in cases with pattern similarity. It was often applied when a numerical threshold (typically eight to twelve minutiae) was required before deciding source attribution.

8.6. In assessing dissimilarities, the examiner must scrutinize the impression concentrating on ridge morphology and specifically considering factors such as distortion, multiple overlapping impressions, compression, and scarring (Figures B2 & B3). Skin creases, which may have diminished or become more prominent over time, could also account for the appearance of misaligned ridges. It is incumbent on the examiner to acknowledge the differences, when present, and fully document the data used in the decision-making process.

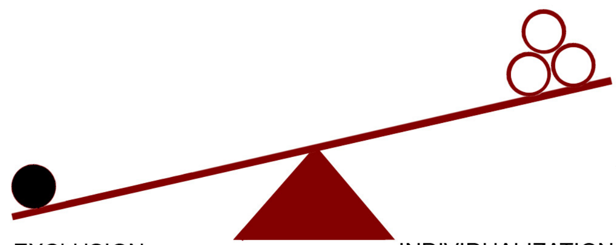


Figure B2: In this instance an exclusion decision has been made because few features supporting individualization and an un-reconciled dissimilarity has been observed and determined by the examiner to be a discrepancy.[7]

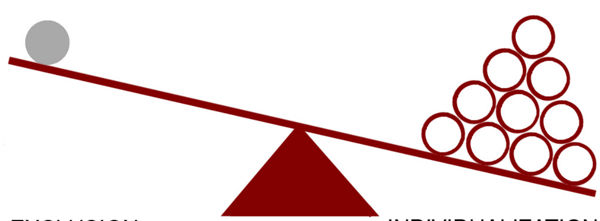


Figure B3: In this instance an individualization conclusion has been reached because the abundance of features supporting individualization outweighs the presence of dissimilarities even though some may be of indeterminate origin.[8]

[7] Illustrations are intended to represent concepts of decision-making and not a numerical standard on which to base those decisions.

8.7. There are instances where, for example, due to accident, disease or intentional disfigurement, fingerprint patterns, or ridge flow may change and take on the appearance of a different pattern or ridge configuration. Examiners making exclusions based on Level 1 detail must be aware of these phenomena and account for their influence as part of the decision making process (Figure B4).

Images courtesy of the Fingerprint Interest Group

Figure B4: The example above illustrates a situation where, due to an injury, the impression on the right has a substantially different Level 1 pattern appearance from the pre-injury impression on the left.

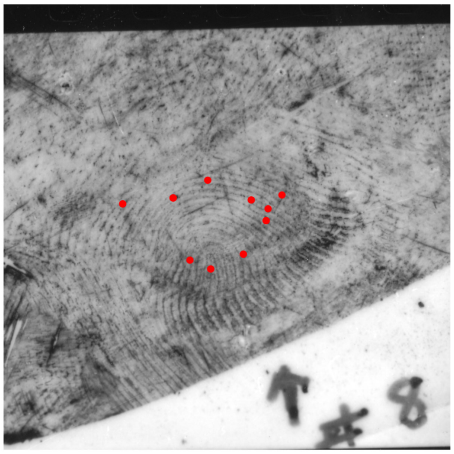
8.8. Extra ridges may appear in a friction ridge impression even though there is an abundance of corresponding features between two impressions. This may be caused by movement of the friction ridge skin or surface at the time of deposition. Examiners making exclusions based on Level 2 detail must be aware of this phenomenon and account for its influence as part of the decision making process (Figure B5).

[8] Illustrations are intended to represent concepts of decision-making and not a numerical standard on which to base those decisions.

Figure B5: In this example, extra ridges, attributed to slippage during recording, resulted in a double tap solely in that area, while the remaining friction ridge surface remained as a single touch.

8.9. When the impressions are complex, the decision requires more expertise, consideration, and documentation of factors such as distortion and multiple depositions. Friction ridge details are not always accurately replicated in low quality impressions and variations in appearance of ridge characteristics and their sequence may be present. It should be noted that variations may also appear in high quality prints. Caution is warranted in the determination of whether a variation is a dissimilarity rather than a discrepancy (Figures B6 and B7).

Figure B6: In the example above there is the potential of an erroneous exclusion based on Level 1 pattern. The impression contains red flags that would be cause to question the reliability of Level 1. Level 2 details should be considered prior to the exclusion conclusion.



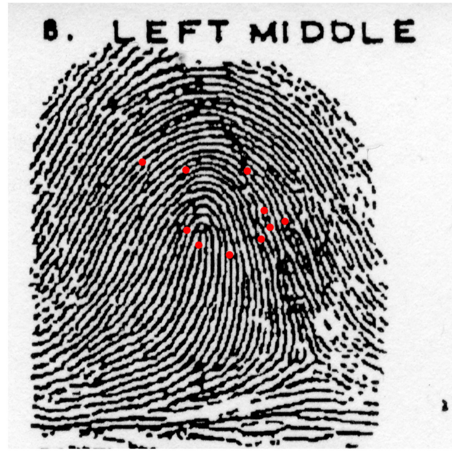


Figure B7: In the real case example above, multiple examiners erroneously excluded the source due to apparent Level 1 ridge flow difference.

9. Conditions Supporting Exclusion

9.1. Expert consensus and experience have determined that source exclusion may be reliably made using Level 1 and Level 2 details. Exclusion relying predominantly on Level 3 detail is problematic due to inconsistency in its recording.

9.2. In order to reach a decision to exclude a source, the examiner must weigh the discriminating value of the features present in the impression that support individualization against those which support exclusion.

10. References

[1] "One-Dissimilarity Doctrine in Fingerprint Identification," J.I. Thornton, International Criminal Police Review, March 1977, Issue 306 Pages 89-95.

[2] SWGFAST Document #19 (2013), Standard Terminology of Friction Ridge Examination (Latent/Tenprint), Ver. 4.0.

11. Revision Table

Version	Effective Start	Effective End	Posted	Archived	Change
1.0	3/15/11	09/13/11		09/13/11	Original Issue
1.1	09/13/11	N/A	10/26/11	N/A	1. Revised numbering 2. Updated Appendix with final reference source version
2.0	03/13/13	N/A	04/27/13	N/A	Added Appendix B Reformatted (start of new version number)

“Information Paper,” Department of the Army , Defense Forensic Science Center, CIFS-FSL-LP (3 November 2015). Document in public domain.

DEPARTMENT OF THE ARMY
DEFENSE FORENSIC SCIENCE CENTER
4930 N 31ST STREET
FOREST PARK, GA 30297

CIFS-FSL-LP 03 November 2015

INFORMATION PAPER

SUBJECT: Use of the term “Identification” in Latent Print Technical Reports

1. Forensic science laboratories routinely use the terms “identification” or “individualization” in technical reports and expert witness testimony to express the association of an item of evidence to a specific known source. Over the last several years, there has been growing debate among the scientific and legal communities regarding the use of such terms within the pattern evidence disciplines to express source associations which rely on expert interpretation. Central to the debate is that these terms imply absolute certainty of the conclusion to the fact-finder which has not been demonstrated by available scientific data. As a result, several well respected and authoritative scientific committees and organizations have recommended forensic science laboratories not report or testify, directly or by implication, to a source attribution to the exclusion of all others in the world or to assert 100% certainty and state conclusions in absolute terms when dealing with population issues.

2. The Defense Forensic Science Center (DFSC) recognizes the importance of ensuring forensic science results are reported to the fact-finder in a manner which appropriately conveys the strength of the evidence, yet also acknowledges that absolute certainty should not be claimed based on currently available scientific data. As a result, the DFSC has modified the language which is used to express “identification” results on latent print technical reports. The revised languages is as follows:

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

3. This revision to the reporting language is not the result of changes in the examination methods and does not impact the strength of the source associations. Instead, it simply reflects a more scientifically appropriate framework for expressing source associations made when evaluating latent print evidence. The next step will be to quantify both the amount of corresponding ridge detail and the related likelihood calculations. In the interim, customers should continue to maintain strong confidence in latent print examination results.

4. References:

a. National Research Council (2009). Strengthening Forensic Science in the United States: A Path Forward. National Research Council, Committee on Identifying the Needs of the Forensic Science Community. National Academies Press, Washington, D.C.

b. National Institute of Standards and Technology (2012). Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach. Expert Working Group on Human Factors in Latent Print Analysis, U.S. Department of Commerce, National Institute of Standards and Technology.

c. Garrett, R. (2009). Letter to All Members of the International Association for Identification, Feb. 19, 2009.

5. Questions regarding this information paper may be directed to Mr. Henry Swofford, Chief, Latent Print Branch, USACIL, DFSC, 404-469-5611 and Henry.J.Swofford.Civ@mail.mil.

CLASS 5

William C. Thompson, "The National Research Council's Plan to Strengthen Forensic Science: Does the Path Forward Run Through the Courts" 50 Jurimetrics 35-51 (Fall 2009), online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2214475.

Roger Koppl, "How to Improve Forensic Science" European Journal of Law and Economics, 20: 255-286 (2005), online at: http://inside.fdu.cdu.edu/pt/koppl.html.

School of Law
University of California • Irvine

Legal Studies Research Paper Series No. 2011-22

Fordham Urban Law Journal, Vol. 38, No. 2, 2010

ACCULTURATING FORENSIC SCIENCE: WHAT IS 'SCIENTIFIC CULTURE', AND HOW CAN FORENSIC SCIENCE ADOPT IT?

Simon A. Cole
scole@uci.edu

University of California, Irvine ~ Department of Criminology, Law and Society

ACCULTURATING FORENSIC SCIENCE: WHAT IS 'SCIENTIFIC CULTURE', AND HOW CAN FORENSIC SCIENCE ADOPT IT?

Simon A. Cole[*]

INTRODUCTION

The topic for this special issue is: "How should judges, legislators, and the legal community in general respond to the National Research Council of the National Academy of Sciences' 2009 report, *Strengthening Forensic Science in the United States*?" I think there are some fairly easy answers to this question that should not brook a great deal of controversy or disagree-

[*] Associate Professor of Criminology, Law & Society, University of California, Irvine; Ph.D. (Science & Technology Studies), Cornell University; A.B., Princeton University. For discussion, references, and advice, I am grateful to Douglas Haynes, Sheldon Greenfield, Shobita Parthasarathy, Adina Schwartz, Gary Edmond, Beth Bechky, Roger Koppl, Corrina Kruse, Birgitta Rasmusson, Carroll Seron, Laura Kelly, and Norah Rudin. Thanks also to the student editors at Fordham Law School for their hard work on this Article. Some of my ideas for this Article were developed during a joint research visit to the Swedish National Forensic Laboratory and Linköping University. I am grateful to both institutions for their generosity and hospitality. This material is partially based upon work supported by the National Science Foundation under grant No. SES-0115305. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

ment. First, validation studies[1] should be performed for forensic assays[2] for which they have not yet been performed.[3] Although the NAS Report is fairly clear about the absence of validation studies, some controversy remains over whether validation studies have been performed for some assays. Therefore, Dr. Bohan suggests that "validation investigations" should be performed to assess the state of validation of each assay.[4] These might be followed by validation studies.

Second, the proposed National Institute of Forensic Science (NIFS) should be created.[5] Certainly, there are potential downsides and criticisms, but creating NIFS is probably better than the status quo. If created, the Institute should be in the form proposed by the NAS Report.[6] Crucial aspects of this form include that it should be an independent agency, especially independent of law enforcement, and a proper scientific organization staffed by scientists. If it is "captured"[7] by law enforcement, it becomes less obvious that it would be a force for improvement rather than stagnation.

Third, the reporting of forensic analyses results should be reformed and standardized such that they are scientifically supportable.[8] Judges should restrict the admissibility of forensic assays that lack the aforementioned validation.[9]

Finally, judges, legislators, and the legal community should contemplate the broader meaning of the NAS Report's conclusion that the courts' handling of forensic evidence over the past couple of decades has been "utterly ineffective."[10] Setting forensic evidence aside, what weaknesses in our current system of justice does this "utter ineffectiveness" identify? Re-

---

1. "[A] validation study is designed to measure the accuracy of a scientific technique. The Study attempts to identify and quantify the inherent margin of error in the technique." Edward J. Imwinkelried, *Coming to Grips With Scientific Research in* Daubert*'s "Brave New World": The Courts' Need to Appreciate the Evidentiary Differences Between Validity and Proficiency Studies*, 61 BROOK. L. REV. 1247, 1254 (1995) (internal citations omitted).

2. An "assay" might also be called a "forensic test" or a "forensic technique," and would include, inter alia, tests of techniques involving the comparison of latent prints, firearms and tool marks, handwriting identifications, and bite marks.

3. NAT'L RES. COUNCIL, NAT'L ACADEMY SCI., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 22 (2009) [hereinafter NAS REPORT].

4. Thomas L. Bohan, *Strengthening Forensic Science: A Way Station on the Journey to Justice*, 55 J. FORENSIC SCI. 5, 6 (2010).

5. NAS REPORT, *supra* note 3, at 19.

6. *Id.*

7. Roger Koppl, *Who Will Capture Forensic Science?*, THINKMARKETS (Apr. 6, 2009), http://thinkmarkets.wordpress.com/2009/04/06/who-will-capture-forensic-science/#more-1289.

8. NAS REPORT, *supra* note 3, at 21.

9. *Id.* at 85-86.

10. *Id.* at 109.

forms to the justice system should be enacted so that in the future, courts' handling of scientific issues is less likely to become a glaring embarrassment to the legitimacy of the courts.

These are all things that I think many scholars will agree *should* be done. Whether they are likely to be done and what is the best strategy to assure that they are done will command somewhat less consensus and may well be the subject of other contributions to this special issue. Indeed, in some cases, compelling arguments may be made that half-hearted, ill-conceived attempts to implement some of these reforms may end up making the situation worse rather than better.

All that being said, however, simply asserting that the recommendations of the NAS Report should be followed seems to avoid a much larger issue. An NAS Report is, after all, a highly inefficient, expensive, and slow way of accomplishing tasks. It is also "one-off," in the sense that we cannot realistically expect there to be periodic NAS Reports indicating what needs to be done in forensic science, especially given the well known difficulties there were in bringing this particular NAS Report into being.[11] The larger issue is why the aforementioned actions never occurred in the first place. In other words, why was it necessary for the National Academy of Sciences (NAS) to intervene in 2009 to demand that validation studies be conducted a century after the introduction of latent print evidence into court,[12] nearly eighty years after the introduction of firearms and tool mark evidence,[13] nearly fifty years after the introduction of bite mark evidence,[14] and more than a century after the introduction of handwriting evidence?[15] Why did the NAS Report have to state that forensic science should be treated as science and not as an arm of law enforcement,[16] or call for the reporting of forensic results to be both standardized and scientifically supportable?[17] Why was it necessary for the NAS to suggest that forensic science be better regulated, or that accreditation of laboratories and certification of analysts be mandatory?[18]

---

11. Donald Kennedy, Editorial, *Forensic Science: Oxymoron?*, 302 SCIENCE 1625 (2003).

12. Simon A. Cole, *Grandfathering Evidence: Fingerprint Admissibility Ruling From* Jennings *to* Llera Plaza *and Back Again*, 41 AM. CRIM. L. REV. 1189, 1216-17 (2004).

13. MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman et al. eds., 3d ed. 2007).

14. *Id*.

15. *Id*.

16. NAS REPORT, *supra* note 3, at 19.

17. *Id*. at 21.

18. *Id*. at 23.

If we pose the questions this way, then the answer to the question, "what is to be done?" takes on a different tone. In addition to the short term actions listed above, a long-term structural solution is required to ensure that past mistakes are not repeated regardless of NAS intervention. Again, NAS intervention is an extremely costly, inefficient, time-consuming, and unrealistic mechanism for governing an area of scientific activity. What is wanted is for forensic science to become self-governing in such a way that future NAS Reports become unnecessary and that future outside observers will not find lacunae quite as great as the ones discovered by this NAS Committee (validation, accreditation, certification, standardization of testimony, and so on). The hope then would be that fifty years from now, rather than convening another NAS committee which finds new deficiencies in forensic science (or, worse, finds that the existing "serious deficiencies in the nation's forensic science system"[19] have still not been addressed), those actions that a future NAS committee would expect to have occurred will have occurred naturally.

The NAS Report does not focus very much on the "why" questions, but, to the extent that it does, its discussion is centered around what it calls the "culture of science."[20] The Report describes the "culture of science" as an important missing ingredient in at least parts of forensic science.[21] It states that "some . . . activities" that fall under the broad rubric of "'forensic science' . . . are not informed by scientific knowledge, or are not developed within the culture of science."[22] Further, it touts "scientific culture" as a potential antidote to what one of the Committee co-chairs called elsewhere "The Problems That Plague the Forensic Science Community."[23] "The forensic science disciplines will profit enormously by full adoption of this scientific culture."[24] Moreover, the Report asserts that "a culture that is strongly rooted in science" is a "minimum" criterion for the new federal agency it proposes, the National Institute of Forensic Science (NIFS).[25] It

19. News Release, Nat'l Academies, 'Badly Fragmented' Forensic Science System Needs Overhaul: Evidence to Support Reliability of Many Techniques is Lacking (Feb. 18, 2009), *available at* http://www8.nationalacademies.org/onpinews/newsitem.aspx?RECORDID=12589.

20. NAS REPORT, *supra* note 3, at 39.

21. *Id.*

22. *Id.*

23. *See id.* at 114, 125; Hon. Harry T. Edwards, Senior Circuit Judge and Chief Judge Emeritus, D.C. Cir., Keynote Address at Conference on Forensic Science for the 21st Century: The National Academy of Sciences Report and Beyond, Solving the Problems That Plague the Forensic Science Community (Apr. 3, 2009).

24. NAS REPORT, *supra* note 3, at 125.

25. *Id.* at 18.

would seem, then, that in addition to following the recommendations of the NAS Committee, forensic science should adopt scientific culture, and this Article will focus on that question.

In focusing on this issue, I believe that I have taken on a more difficult task than, arguing, as I have elsewhere,[26] that validation studies of forensic techniques should be conducted or even that evidence should not be admissible without such validation studies.[27] There are vigorous debates about what "validation" means,[28] and since the NIFS has not yet been created, there are disagreements about what "it" is.[29] Even so, validation, the NIFS, and standardized testimonial reporting are concrete things compared to the nebulous notion of "scientific culture." What is this "scientific culture" that the NAS Committee says is both missing and needed in at least some parts of forensic science? The Report never explicitly defines scientific culture. Turning to the scholarly literature is of little help. "Scientific culture" is a rather vague and contested term that is used to mean a variety of different things.[30] If no one agrees upon what we mean by "scientific culture," then the NAS Report's call to "adopt" it becomes an empty rhetorical gesture, easily answered by any interest group that simply chooses its preferred definition of "scientific culture" and declares that forensic science has or has not adopted it. I will argue, however, that all is not lost merely because of the vagueness of "scientific culture" and the NAS Report's discussion of it. To the contrary, it is important, and perhaps indispensable, that we can articulate precise meanings for the term, and that forensic science adopt something called "scientific culture" if any of the commonly desired responses to the NAS Report articulated above are to occur.

I should note that, while I will be critical of the NAS Report's treatment of the notion of "scientific culture," this should not be construed as criticism of the Report as a whole or of the NAS Committee. In my view, the

---

26. *See* Simon A. Cole, *Comment on 'Scientific Validation of Fingerprint Evidence Under* Daubert*,'* 7 LAW PROBABILITY & RISK 119 (2007); Simon A. Cole, *Is Fingerprint Identification Valid? Rhetorics of Reliability in Fingerprint Proponents' Discourse*, 28 LAW & POL'Y 109 (2006).

27. *See* Cole, *supra* note 12.

28. *See, e.g.*, Lyn Haber & Ralph Haber, *Scientific Validation of Fingerprint Evidence Under* Daubert, 7 LAW PROBABILITY & RISK 87, 88 (2008); Imwinkelried, *supra* note 1, at 1256-60.

29. *See* Imwinkelried, *supra* note 1, at 1256-60; Haber & Haber, *supra* note 28.

30. *Compare* Sarah Franklin, *Science as Culture, Cultures of Science*, 24 ANN. REV. ANTHROPOLOGY 163 (1995), *with* Benoit Godin & Yves Gingras, *What Is Scientific and Technological Culture and How Is it Measured? A Multidimensional Model*, 9 PUB. UNDERSTANDING SCI. 43 (2000).

Report's lucid discussion of such issues as validation and "individualization" far outweigh the ambiguities I will identify around the notion of "scientific culture." Indeed, it is partly because I agree so wholeheartedly with so much of what is elsewhere in the Report that I have chosen to engage in this extended attempt at developing further the Report's discussion of "scientific culture."

In Part I, I describe how the NAS Report characterizes "scientific culture." I suggest that the described attributes of scientific culture are vague and unspecific, and that more thought is necessary to elucidate how they might map onto forensic science. In Part II, I suggest that the NAS Report's characterization of "scientific culture" is based on popular accounts of science and "the scientific method." I suggest that these accounts are incomplete, generally considered obsolete, and not particularly helpful in pointing a way toward reform of forensic science. In Part III, I posit a conception of science as *work* rather than *method*. In Part IV, I offer a tentative mapping of how forensic science might be understood as work by dividing forensic labor in a set of general tasks. In Part V, I offer a tentative mapping of the goals and desired attributes of scientific workers who would perform each type of forensic task. In Part VI, I briefly describe how the status quo seems to fall short of the desired situation described in Part V. In Part VII, I suggest that medicine offers a reasonable analogy for the sort of structuring of labor into tasks that might be desirable for forensic science. I conclude with some observations and clarifications about the medical model I proposed.

It should also be noted that my comments about the state of forensic science are, like the NAS Report itself, exclusively concerned with forensic science as it exists in the United States. While many of the issues identified by the NAS Report (i.e., validation) transcend national borders, many others (i.e., standardization) do not. While the NAS Committee was obviously limited in terms of what it could accomplish given its constraints on time and resources, the NAS Report can be reasonably criticized for neglecting to look outside the United States for potential solutions to the problems it identified. Indeed, some of my prescriptive remarks are drawn from my limited knowledge of practices outside the United States.

I. THE NAS REPORT'S TREATMENT OF SCIENTIFIC CULTURE

The NAS Report's discussion of scientific culture is contained almost entirely in Chapter 4, "Principles of Science and Interpreting Scientific Data."[31] In this chapter, the report offers the sort of account of "the scientific

31. NAS REPORT, *supra* note 3, at 111.

method" often written by non-philosophers, such as journalists, lawyers, policy-makers, or scientists themselves.[32] The account draws heavily on Karl Popper's theory of falsifiability.[33] According to the Report, there is a single scientific method called "the scientific method."[34] The method consists of posing hypotheses and measuring them against data, resulting in either refutation or support. "Absolute truth" is never achieved, but this process of continual testing "approaches truth . . . incrementally."[35]

> Acceptance of the work comes as results and theories continue to hold, even under the scrutiny of peers, in an environment that encourages healthy skepticism. That scrutiny might extend to independent reproduction of the results or experiments designed to test the theory under different conditions. As credibility accrues to data and theories, they become accepted as established fact and become the "scaffolding" upon which other investigations are constructed.[36]

The NAS Report also offers an account of the "principles of science" that consists of another common attribute of such accounts of science: a recitation of supposed virtues common to scientists and the scientific way of thinking.[37] A large body of thought has historically attributed the material and epistemological success of science to a particular virtuous way of thinking. Such arguments have a long history dating back at least as far as Sir Francis Bacon.[38] In their most modern form, however, they tend to be associated with the sociologist Robert Merton, who famously articulated four "norms," associated with the profession of science, although Merton treated these attributes as social norms, not epistemological virtues.[39] Merton's four norms were: communism (collective ownership of data and knowledge), universalism (scientific knowledge is "impersonal," in the sense that its truth value has nothing to do with the personal attributes of the individuals who develop it), disinterestedness (lack of interest in any particular outcome of the search for scientific truth), and organized skepticism (the collective subjection of all ideas to the highest level of scrutiny).[40]

The NAS Report implicitly invokes all four of Merton's norms in its description of how "scientific culture" ought to function. Theories and data

32. NAS REPORT, *supra* note 3, at 112.
33. KARL R. POPPER, THE LOGIC OF SCIENTIFIC DISCOVERY 57-74 (1959).
34. NAS REPORT, *supra* note 3, at 112.
35. *Id.*
36. *Id.*
37. *Id.*
38. *See generally* FRANCIS BACON, THE NEW ORGANON (2000).
39. *See* ROBERT K. MERTON, SOCIAL THEORY AND SOCIAL STRUCTURE: TOWARD THE CODIFICATION OF THEORY AND RESEARCH (1949).
40. *Id.*

should be shared through "collegial interactions" (communism).[41] "The need for credibility among peers drives investigators to avoid conflicts of interest" (disinterestedness).[42] Science should be characterized by "openness to new ideas."[43] "The scientific culture encourages continued questioning and improvement."[44] "A scientist encounters . . . unconscious bias if he/she becomes too wedded to a preliminary conclusion, so that it becomes difficult to accept new information fairly and unduly difficult to conclude that the initial hypotheses were wrong."[45] Science should be "a self-correcting enterprise."[46]

> Science is characterized also by a culture that encourages and rewards critical questioning of past results and of colleagues [organized skepticism]. . . . The scientific culture encourages cautious, precise statements and discourages statements that go beyond established facts; it is acceptable for colleagues to challenge one another, even if the challenger is more junior [universalism].[47]

Anecdotal arguments may be made that forensic science, as currently constituted in the United States, demonstrates broad failures of all four norms. Communism is violated by refusals to share data, both broad validation data and specific data from specific cases requested through discovery.[48] Universalism is violated by the tendency toward *ad hominem* attacks on those who challenge the conventional wisdom and the extreme hostility expressed toward outsiders.[49] Disinterestedness is violated by the location

---

41. NAS REPORT, *supra* note 3, at 112.

42. *Id.*

43. *Id.* at 113.

44. *Id.* at 114.

45. *Id.* at 123-24.

46. *Id.* at 125.

47. *Id.*

48. *See, e.g.*, D. Michael Risinger & Michael J. Saks, *A House With No Foundation*, 20 ISSUES IN SCI. & TECH. 35 (2003); William C. Thompson, *Evaluating the Admissibility of New Genetic Identification Tests: Lessons From the 'DNA War,'* 84 J. CRIM. L. & CRIMINOLOGY 22, 99 (1993).

49. *See, e.g.*, David L. Grieve, *Simon Says*, 51 J. FORENSIC IDENTIFICATION 85 (2001); André Moenssens, *Fingerprint Identification: A Valid Reliable "Forensic Science"?*, 18 CRIM. JUST. 31 (2003); André Moenssens, *Handwriting Identification Evidence in the Post-*Daubert *World*, 66 UMKC L. REV. 251 (1997); Leslie Roberts, *Prosecutor v. Scientist: A Cat and Mouse Relationship*, 257 SCIENCE 733 (1992); Thompson, *supra* note 48, at 100-02; Gina Kolata, *Critic of 'Genetic Fingerprinting' Tests Tells of Pressure to Withdraw Paper*, N.Y. TIMES, Dec. 20, 1991, at A20; Ronald Nichols, *The Scientific Foundation of Firearms and Tool Mark Identification—A Response to Recent Challenges*, CAL. ASS'N CRIMINALISTS NEWS (2006), at 8-27; Glenn Langenburg, *Defending Against the Critic's Curse,* DETAIL (Sept. 30, 2002), http://www.clpex.com/Articles/CriticsCurse.htm; André Moenssens, *The Reliability of Fingerprint Identification: A Case Report* (2002), http://www.forensic

of most forensic science in law enforcement agencies associated with a particular "side" of the criminal trial.[50] Organized skepticism is violated by the strong resistance to challenging cherished assumptions.[51]

I would like to note that we could also find similar anecdotes about mainstream science. More importantly, even in "normal" scientific practice, Merton's norms appear to function more as ideals to which to aspire than as determinants of actual behavior. Indeed, one particularly well known empirical study of a set of actual (normal, respectable, non-deviant) scientists involved in a scientific controversy found that, rather than behaving according to Merton's norms, they behaved in precisely the opposite fashion, which were referred to as "counter-norms."[52] These counter-norms were not viewed as deviant behavior on the part of the scientists. Interview data showed that most scientists expected one another to behave in this manner.[53] For example, the notion that real scientists would adhere to dispassion, rather than function as advocates for their chosen theories was regarded by most interview subjects as hopelessly naïve.[54] The purpose of this discussion is not to claim that forensic science is "pathological science."[55] Rather, the point is to draw attention to the extent to which violations of at least some of Merton's norms appear to be openly and deliberately espoused, rather than covertly and inadvertently practiced, and even embedded in the "culture" of forensic science.

The NAS Report's recitation of Popper and Merton is a familiar enough attribute of lay accounts of what science is and how scientists should be-

---

evidence.com/site/ID/pollak2002.html (last visited Nov. 22, 2010); or almost anything published in CRIME LAB REPORT, http://www.crimelabreport.com/.

50. Paul Giannelli, *The Abuse of Scientific Evidence in Criminal Cases: The Need for Independent Crime Laboratories*, 4 VA. J. SOC. POL'Y & L. 439 (1997).

51. *See, e.g.*, DAVID R. ASHBAUGH, QUANTITATIVE-QUALITATIVE FRICTION RIDGE ANALYSIS: AN INTRODUCTION TO BASIC AND ADVANCED RIDGEOLOGY 4 (1999) ("In the past, the friction ridge identification science has been akin to a divine following. Challenges were considered heresy and challengers frequently were accused of chipping at the foundation of the science unnecessarily. This cultish demeanor was fostered by a general deficiency of scientific knowledge, understanding, and self-confidence within the ranks of identification specialists."); *see also* David L. Grieve, *Possession of Truth*, 46 J. FORENSIC IDENTIFICATION 521, 528 (1996) ("[T]he assumption of absolute certainty as the only possible conclusion has been maintained by a system of societal indoctrination, not reason, and has achieved such ritualistic sanctity that even mild suggestions that its premise should be re-examined are instantly regarded as acts of blasphemy.").

52. Ian I. Mitroff, *Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists*, 39 AM. SOC. REV. 579, 587-89 (1974).

53. *Id.* at 589.

54. *Id.*

55. Irving Langmuir, *Pathological Science*, 42 PHYSICS TODAY 36 (1989).

have. But the mere recitation of these epistemological virtues is not enough. How are hypothesis testing and organized skepticism, for example, to be operationalized in everyday forensic practice? That is not clear. In the next section, I will argue that the reason it is not clear is because the NAS Report, like many accounts of science, equates "science" with what I will call "discovery science," while paying little attention to large swaths of scientific practice that cannot be characterized as discovery.

II. SCIENCE AND SCIENTIFIC METHOD

The lack of precision around the term "scientific culture" is closely related to the lack of precision around other honorific terms, such as "science" and "the scientific method."[56] I will argue that there are definitional confusions around all three terms. Moreover, I will suggest that, for all three terms, these confusions are exacerbated by a tendency to equate "science" with what I will call "discovery science"—scientific activity designed to create new, generalizable knowledge about the natural world. In fact, "science" is a far more rich and more varied enterprise than just "discovery science." There are a wide variety of activities that we conventionally consider "science" that would not be described as "discovery science." These activities would include: routine laboratory work, such as work that may be part of larger projects that may themselves be "discovery science"; industrial science; engineering; much of medicine; descriptive science; and so on. There are armies of workers engaged in what sociologists call "technoscientific work" in the enterprise of modern science, who are performing work that could not, in itself, be characterized as discovery.[57] They are manipulating cell cultures, synthesizing molecules, tuning detectors, analyzing samples, and so on.

When lay people, including lawyers and judges, write about "science," however, they tend to write exclusively about "discovery science."[58] This is perhaps because those who first produced public "accounts" of science—philosophers, historians, science journalists, and scientists themselves—were primarily interested in the making of new knowledge about the natural world.[59] Philosophers and historians of science, for example, were interested in how new scientific knowledge was made, not, for example, in what epistemological basis a laboratory technician had for performing a

56. Gary Edmond & David Mercer, *Trashing 'Junk Science'*, 1998 STAN. TECH. L. REV. 3, 28 (1998).

57. *See generally* BETWEEN CRAFT AND SCIENCE: TECHNICAL WORK IN U.S. SETTINGS (Stephen R. Barley & Julian E. Orr eds., 1997).

58. *E.g.*, KARL R. POPPER, THE LOGIC OF SCIENTIFIC DISCOVERY (1959).

59. *Id.*

routine assay. Likewise, the educated public that consumed science journalism and memoirs of scientists were interested in discovery, not mundane laboratory work. Among lay people, these remain the dominant accounts of science.[60] Therefore, when lay people, including judges and attorneys, write or talk about science, they tend to write or talk about discovery science, and they tend to use as authorities accounts that primarily focus on discovery science.[61]

The development of sociology and anthropology of science, however, brought attention to hitherto ignored areas of "mundane" scientific work. Sociologists and anthropologists are not only interested in discovery but also in what scientific practice was actually like *as work*.[62] They focused on what scientists actually did all day, rather than merely isolating those rare moments in which scientists discovered new knowledge.[63]

This is not to argue that sociology and anthropology are better than philosophy and history or vice versa. Rather, it is to argue that there are a variety of activities that encompass the enterprise we call "science" and there are a variety of accounts of these different activities. Lay people, like judges and lawyers, tend to be familiar only with accounts of discovery science and thus seek to apply those accounts to all scientific activities.[64] This results in a mismatch, which can produce a variety of misunderstandings. Mundane scientific activity can be characterized according to explanations that were constructed for discovery science and can be found wanting because it does not meet expectations that were devised for discovery science. Normatively, policy-makers may apply norms that may be appropriate for discovery science to mundane scientific activity. I will suggest that we cannot coherently apply such notions as "science," "scientific method," and "scientific culture" to forensic science without thinking seriously about what sort of scientific activity forensic science purports to be.

---

60. Matthias Nuckles et al., *How Do Experts Adapt Their Explanations to a Layperson's Knowledge in Asynchronous Communication? An Experimental Study*, 16 USER MODELING & USER ADAPTED INTERACTION 2 (2006).

61. *See* Gary Edmond & David Mercer, *Conjectures and Exhumations: Citations of History, Philosophy and Sociology of Science in US Federal Courts*, 14 LAW & LITERATURE 309 (2002) [hereinafter Edmond & Mercer, *Conjectures and Exhumations*]; Edmond & Mercer, *supra* note 56.

62. *E.g.*, ANDREW PICKERING, SCIENCE AS PRACTICE AND CULTURE (1992).

63. *Id.*

64. Edmond & Mercer, *supra* note 56.

A. Science

The critique of forensic science found in the NAS Report is sometimes characterized in the popular media as a claim that forensic science is "unscientific."[65] That is not, in fact, a claim made in the NAS Report. Nowhere does it say that forensic science is "not science." Instead, the NAS opted for much more specific claims, such as that forensic science lacks adequate validation, certification, accreditation, oversight, and basic research, among other things.[66] The same is true of critiques published by forensic science reformers prior to the NAS Report.[67] These critiques, likewise, tended to focus on more specific issues than the broad-brush claim that forensic science was "not science."[68]

There are good reasons for forensic science reformers to avoid making the charge that forensic science is "not science." Supporting such a claim would require defining what science is and showing that all of forensic science falls outside that definition. Defining "science" and distinguishing it from "pseudo-science" is a problem that philosophers of science call "demarcation."[69] Current philosophy of science views the demarcation problem as unsolved—that is, there is no single definition of "science" that neatly divides everything upon which we want to bestow the title "science" from everything upon which we don't want to bestow that title.[70] The best known purported "solution" to the demarcation problem, Karl Popper's notion of "falsification,"[71] is not viewed by most contemporary philosophers of science as a complete solution: there are areas of study that we generally consider "science" (descriptive biology, geology, etc.) that do not meet the criteria of "falsifiability."[72] If there is no agreed upon demarcation criterion that neatly divides knowledge claims into "scientific" and "non-scientific" categories in a satisfactory way, then there is little to be gained from arguing about how such a demarcation criterion would apply to forensic science.

65. *See, e.g.*, Yamil Berard, *Stakes are High as Doubt Is Cast on Forensic Lab Techniques*, Fort Worth Star-Telegram, Nov. 28, 2009, *available at* http://www.mcclatchydc.com/2009/11/29/79698/stakes-are-high-as-doubt-is-cast.html.

66. NAS REPORT, *supra* note 3, at 12.

67. *E.g.*, Haber & Haber, *supra* note 28.

68. *Id.*

69. POPPER, *supra* note 58, at 11.

70. A.F. CHALMERS, WHAT IS THIS THING CALLED SCIENCE? (1999).

71. POPPER, *supra* note 58, at 18 (this is the notion that you can never prove a theory true, but you can prove it false by a contravening example).

72. *Id.*; *see also* CHALMERS, *supra* note 70.

A second reason that there is little use in debating whether or not forensic science is "science," is that it is legally irrelevant. In *Kumho Tire v. Carmichael*, the United States Supreme Court, perhaps in part out of recognition of the problems with demarcation described in the preceding paragraph, relieved courts of the responsibility to decide whether various forms of expert evidence should constitute "science" or "non-science."[73] The Court ruled that the same criteria enumerated in its *Daubert v. Merrell Dow Pharmaceuticals, Inc.* decision applied equally to all forms of expert evidence.[74]

The *Daubert* ruling invoked "falsificationism," which has led some to think that the Court adopted it as a demarcation criterion.[75] The Court, however, considerably leavened its invocation of falsification with discussions of alternative, some would say incompatible, markers of "science."[76] The fabled "incoherence" of the *Daubert* decision has occasioned a great deal of criticism from scholars.[77] We need not rehash all of this criticism here; however, one issue is particularly pertinent to our discussion. In turning to philosophy of science in an attempt to articulate a definition of reliable science, the *Daubert* Court necessarily turned to a literature that is concerned with what we might call "discovery science."[78] Philosophers, like those cited prominently in the *Daubert* decision, were concerned with the sort of science practiced by scientists who are trying to develop new knowledge about the natural world (e.g., discover the laws of physics, understand the evolution of species, etc.).[79] While this sort of activity captures the attention of philosophers, there is a great deal of activity upon which we would bestow the title of "science" that is nothing like "discovery science." For example, laboratory technicians performing routine assays, industrial scientists seeking to refine a product or process, and even physicians trying to diagnose patients or engineers trying to design a safer bridge might defensibly be called "science," and yet are probably not best described as efforts to discover generalizable truths about the natural world.

The *Daubert* decision is generally believed to have been occasioned, in part, by a perception that that the courts were facing an increasing number

73. Kumho Tire v. Carmichael, 526 U.S. 137, 147-49 (1999).

74. *See* Daubert v. Merrell Dow Pharm. Inc., 509 U.S. 579, 592-94 (1993).

75. Susan Haack, *An Epistemologist in the Bramble-Bush: At the Supreme Court With Mr. Joiner*, 26 J. HEALTH POL. POL'Y & L. 217 (2001).

76. *Id.* at 226.

77. *Id.* at 232-33.

78. *See* Edmond & Mercer, *Conjectures and Exhumations*, *supra* note 61, at 310-12 (2002); *cf.* POPPER, *supra* note 58.

79. POPPER, *supra* note 58, at 37-38.

of highly technical scientific issues.[80] *Daubert*, in this formulation, was an effort by the Court to assist the lower courts by providing guidance in dealing with such issues.[81] But most legal issues involving science do not seem to involve discovery science. There are few legal cases that hinge on, say, particle physics or string theory. Instead, most legal issues would seem to pertain to precisely those "other," more mundane kinds of science: laboratory tests, medical diagnoses, epidemiology, and industrial science, among others. The paradox is that, in seeking to assist lower courts in dealing with a perceived flood of scientific issues involving what we might call "mundane science," the Court drew on a literature constructed exclusively around the philosophical problems raised by discovery science. This mismatch has not been sufficiently recognized, either by the courts themselves or in the scholarly literature, but it surely is behind much of the sense of dissatisfaction that continues to surround the *Daubert* regime. This mismatch will be crucial to my argument in this Article.

B. Scientific Method

While "science" may be too broad a concept to define precisely, it is sometimes argued that the "scientific method" might be a somewhat narrower, and thus more precisely definable, concept.[82] In defining the "scientific method, however, we run into the same problem we encountered in seeking to demarcate "science." Popular parlance may believe that there is a single unitary "scientific method" involving the same experimentation and hypothesis testing that all areas of science utilize. Philosophers of science, however, are in broad agreement that there is no single method employed by all areas of knowledge production that we generally call science.[83] As one philosopher of science sums it up, "there is no scientific method."[84] There is, therefore, very little purpose to arguing about whether forensic science uses the "scientific method" if not all areas of conventional science use that method.

It should also be noted, however, that discussions of "scientific method" also suffer from the same mismatch between "discovery science" and

---

80. *See generally* PETER W. HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM (1991).

81. JUSTICE STEPHEN BREYER, FED. JUDICIAL CTR., INTRODUCTION, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 2 (2000) ("The legal disputes before us increasingly involve the principles and tools of science.").

82. *See infra* note 87.

83. Susan Haack, *Not Cynicism, but Synechism: Lessons From Classical Pragmatism*, 41 TRANSACTIONS CHARLES S. PEIRCE SOC'Y 239, 249 (2005).

84. *Id.*

"mundane science." The mythical "scientific method" that looms so large in popular conceptions is associated with the Scientific Revolution and classical experiments designed to discover new facts about the natural world.[85] There is no reason that the method that Galileo or Newton used to develop new knowledge about physics should necessarily be used by a technician working in a molecular biology laboratory, a physician making a diagnosis, an industrial chemist seeking to improve a product or process, an engineer seeking to design a bridge, or a forensic scientist seeking to shed light on a criminal case. The fact that such scientific workers do not use "the scientific method" does not render what they do "not science," or, even more importantly, wrong.

Nonetheless, there appears to be a widespread misconception that any activity that deserves to call itself "science" should be able to account for itself in terms of the "scientific method" of hypothesis testing.[86] This misconception has led to a number of recent efforts to shoehorn forensic science into this canonical "scientific method."[87] These works all characterize the "scientific method" as a multi-step process focused on hypothesis testing. Consider, e.g., several efforts to argue that "ACE-V," the acronym latent print examiners use to describe their supposed "method,"[88] is "synonymous" with "the scientific method." The motivation to associate ACE-V with the honorific "scientific method" is clear: the "scientific method" and hypothesis testing supposedly "ensure[] that the conclusions are the best conclusions possible, given the available data" and make it "certain that the results are sound and supported."[89] Best of all, "[t]he overwhelming popularity of hypothesis testing is due to the reliability of the results when the process is used correctly."[90] Such arguments seem to take the following form:

A. Use of the "scientific method" produces knowledge that is good, valid, and true.

B. Forensic assay X follows the scientific method.

C. Forensic assay X produces results that are good, valid, and true.

---

85. CHALMERS, *supra* note 70.

86. *See infra* note 87.

87. Mark A. Acree, *What Is Science? The Dilemma of Fingerprint Science Revisited*, 14 PRINT 4 (1998); Michelle Reznicek et al., *ACE-V and the Scientific Method*, 60 J. FORENSIC IDENTIFICATION 87, 88 (2010); Michele Triplett & Lauren Cooney, *The Etiology of ACE-V and its Proper Use: An Exploration of the Relationship Between ACE-V and the Scientific Method of Hypothesis Testing*, 56 J. FORENSIC IDENTIFICATION 345, 345 (2006); Pat A. Wertheim, *Scientific Comparison and Identification of Fingerprint Evidence*, 16 PRINT 4 (2000).

88. Sandy L. Zabell, *Fingerprint Evidence*, 13 J.L. & POL'Y 143, 154 (2005).

89. Triplett & Cooney, *supra* note 87, at 346.

90. *Id.*

Both the premises and conclusions of this syllogism are wrong. It is not claimed that hypothesis testing necessarily produces true knowledge. Lots of statements have high truth value without using the "scientific method." In short, it is neither necessary nor sufficient to show that latent print identification is hypothesis testing in order to show that it is valid or reliable. Even setting this point aside, however, these efforts seem misguided. In addition to some more minor misstatements, all of these models suffer from the flaw of conceptualizing the individual case as the hypothesis, rather than the validity of the overall assay. In other words, they conceptualize the null hypothesis as "the unknown friction ridge impressions come from the same source as those of the known print."[91] This hypothesis is then supposedly tested by a comparison of details until the examiner decides that the null hypothesis can be rejected. This conclusion is then supposedly "confirmed" by the replication of this process by another examiner.

But this is not the question that scholars, lawyers, judges, and the NAS have been asking about latent print identification. The question that they have been asking is, "Is latent print identification valid?" Here the hypothesis would be that ACE-V correctly discriminates the true source of latent prints with X degree of accuracy. While not everything that we consider "science" is subject to hypothesis testing, *this* hypothesis is actually susceptible to experimental testing. Such experiments, however, have not been conducted.[92] Therefore, hypothesis testing is reconfigured as a process that applies to individual cases. This is a distortion of the notion of hypothesis testing as classically understood, which is aimed at testing "a general law, rather than a singular statement."[93]

The misuse of this notion of hypothesis testing is demonstrated by the fact that the analogy chosen by the most recent of these articles concerns plant identification.[94] The authors suggest that the classification of an individual sample of plant as a particular species is an example of the "scientif-

91. Reznicek et al., *supra* note 87, at 96.

92. *See* Haber & Haber, *supra* note 28; Simon A. Cole, *Is Fingerprint Identification Valid? Rhetorics of Reliability in Fingerprint Proponents' Discourse*, 28 LAW & POL'Y 109 (2006).

93. HILARY PUTNAM, MATHEMATICS, MATTER AND METHOD 265 (1979). It has been suggested that specific statements may be amenable to a specific type of hypothesis testing, through a Bayesian likelihood ratio approach. Although I am not sure that assessing the relative probabilities of competing hypotheses through a Bayesian approach should necessarily be characterized as "hypothesis testing," it is sufficient here to note that none of the articles discussed *supra* adopt a Bayesian approach. Rather, they are conceptualizing "hypothesis testing" in an "old-fashioned" Popperian sense in which hypotheses are tested through "crucial," potentially falsifying "experiments," in which each successive observation is conceptualized as an "experiment."

94. *See* Reznicek, *supra* note 87, at 87.

ic method" and "hypothesis testing."[95] This is incorrect. The identification of an individual sample of plant as a particular species would be more appropriately characterized as the application of an accepted set of heuristics. The generalizable hypothesis about the natural world in this area of science would be something much broader, like "application of this set of taxonomic rules produces correct plant identifications at X rate." In short, these authors have confused the task of analysis with the task of discovery research, and they treat an analytic activity as if it were an experimental activity.

These efforts seem misguided, not merely because they make erroneous assertions about the "scientific method," and not merely because they seem to distort that "method" in their efforts to make it "fit" forensic science, but, more importantly, because they are fundamentally unnecessary. To answer the questions raised by the NAS Report, it is neither necessary nor sufficient to show that forensic science uses the mythical "scientific method." Showing that forensic science uses "the scientific method" does not answer the NAS Report's claims that it lacks validation, certification, accreditation, oversight, and basic research, nor does it answer the simpler, and yet crucially important question: how accurate is it? Likewise, forensic science could show that it *does* have validation, certification, accreditation, oversight, and basic research without showing that it uses the "scientific method." Importantly, method is unrelated to validity and accuracy. There could be forensic assays that use the "scientific method" in the way that is laid out by these articles, and yet have very low accuracy, resulting in the seemingly paradoxical conclusion that the technique uses the "scientific method" and is usually wrong. Likewise, there could be techniques, perhaps some forensic techniques, that do not use the "scientific method" but are highly accurate.

III. SCIENCE AS WORK

Scholars have already made the point that lawyers and judges tend to invoke philosophical and sociological models of science that, among professional philosophers and sociologists of science, would be viewed as obsolete.[96] Popper's falsificationism has been widely discarded as, at best, a normative description of how science *should* operate, rather than a histori-

95. *Id.*

96. David S. Caudill & Lewis H. LaRue, NO MAGIC WAND: THE IDEALIZATION OF SCIENCE IN LAW 3 (2006); David S. Caudill & Richard E. Redding, *Junk Philosophy of Science?: The Paradox of Expertise and Interdisciplinarity in Federal Courts*, 57 WASH. & LEE L. REV. 685 (2000); Edmond & Mercer, *Conjectures and Exhumations*, *supra* note 61; Gary Edmond & David Mercer, *The Invisible Branch: The Authority of Science Studies in Expert Evidence Jurisprudence*, *in* EXPERTISE IN REGULATION AND LAW 197 (2004).

cally or sociologically accurate model of how it, in fact, does operate.[97] Even more importantly, most contemporary philosophers view falsificationism as inadequate to account for the material and epistemological success of modern science.[98] Likewise, Merton's norms have been widely discarded as idealized notions of how science should operate.[99]

My point here, however, is not merely to take the NAS Committee to task for invoking obsolete models of science. In and of itself, this invocation does no real harm. Rather I want to explore whether there is indeed any harm wrought by these obsolete models of science. In other words, in what ways does the NAS Report's model of science impede the achievement of its purported goal: the "adoption," by forensic science, of a vaguely articulated "scientific culture"?

The most obvious problem is, once again, the mismatch between "discovery science" and "mundane science." Popper's theory of falsificationism was developed in order to explain how scientists made discoveries about the natural world. One aspect of falsificationism that makes this particularly clear—an aspect tellingly missing from the NAS Report's discussion of the "scientific method"—concerns the notion of "boldness." Popper argued that theories should be "bold," that only by thinking big, taking risks, and making "bold conjectures" would scientists advance knowledge.[100] Popper viewed "boldness" as harmless to the enterprise of science (if not to the reputation of the individual scientist) because, he claimed, the rest of the scientific community was busily engaged in attempts at refutation.[101] Thus, bold false hypotheses would be quickly refuted.[102]

It should be clear from this, however, that Popper's theory applies to the sort of theory-generating scientist who works at the apex of the academic establishment. We want boldness from these scientists, but we do not desire boldness from a host of other scientific workers. We do not, as a society, necessarily desire boldness from laboratory technicians, civil engineers, primary care physicians, environmental engineers, public health scientists, and so on. For that matter, we probably do not desire boldness from most forensic scientists. As will be discussed *infra*, most technoscientific work-

97. CHALMERS, *supra* note 70.

98. *Id.*

99. Mitroff, *supra* note 52.

100. *See generally* KARL POPPER, CONJECTURES AND REFUTATIONS (1965).

101. *Id.*

102. Hence the title of one of Popper's books, *Conjectures and Refutations*. *Id.* There is a scientific adage that while bad data is harmful, there is nothing harmful about a bad theory.

ers, known as "forensic scientists," fall into the categories of technoscientific workers from whom we would probably *not* desire boldness.

Likewise, Merton's norms were designed to apply to scientists engaged in open-ended discovery activities. It is far from clear that the paramount virtues for laboratory technicians, civil engineers, primary care physicians, environmental engineers, and public health scientists are Merton's norms. For these categories of scientific workers, it is possible that other virtues, like adherence to procedures and "good hands" are more valuable virtues.

The NAS Report, therefore, uses a philosophy of science designed for discovery scientists to lay out normative goals for a field that is primarily populated by individuals who are not engaged in discovery science. I will argue that this is a serious problem in that it turns the urging that forensic science "adopt scientific culture" into empty sloganeering, rather than a practical set of measures appropriate to the sort of work most forensic scientists do.

Paradoxically, the NAS Report explicitly recognized this mismatch. The report acknowledges that the methods, principles, and virtues it touts "have been discussed here in the context of creating new methods and knowledge,"[103] rather than in the context of routine, quotidian laboratory work. The NAS Report, however, immediately dismisses any concerns about this mismatch in the very same sentence, declaring that

> the same principles hold when applying known processes or knowledge. In day-to-day forensic science work, the process of formulating and testing hypotheses is replaced with the careful preparation and analysis of samples and the interpretation of results. But that applied work, if done well, still exhibits the same hallmarks of basic science: the use of validated methods and care in following their protocols; the development of careful and adequate documentation; the avoidance of biases; and interpretation conducted within the constraints of what the science will allow.[104]

Did the eminent scientists on the NAS Committee really believe that formulating hypotheses is analogous to preparing samples? Does one need to apply, or even understand, falsification in order to prepare a sample? Does one need to adhere to organized skepticism in order to prepare a sample? Did they really believe that the methods, principles, and virtues necessary to use a validated method are the same as those necessary to develop or validate a method? Or did the committee, lacking ready access to a philosophy and sociology of laboratory *work* that is as easily digestible as

103. NAS REPORT, *supra* note 3, at 113.

104. *Id*.

Popper and Merton's accounts of scientific discovery, simply decide to graft accounts of scientific discovery onto scientific work and hope for the best?

The irony, of course, is that plenty of sociological literature on scientific work is available.[105] Again, my purpose here is not to engage in the rather empty exercise of criticizing the NAS Committee for paying inadequate attention to the sociology of science. The NAS Committee had a broad ambit, and it did a fantastic job with many important matters. As a sociologist of science myself, I do not believe that the Committee's limited time would have been well spent parsing the intricacies of the sociology of scientific and technical work. I do believe, however, that the NAS Report's rather cursory engagement with the nature of forensic scientific work demands that others—and who better than sociologists of science?—think more carefully about what "adopting scientific culture" might mean, given the nature of forensic scientific work. I also believe that, if such careful thinking is not done, the NAS Report's call to adopt scientific culture risks becoming an empty gesture.

IV. FORENSIC WORK

I have argued that it is important to understand the practice of science as *work*, and I have argued that what we call "scientific work" encompasses a wide variety of different tasks. How might we begin to describe forensic science as work? We might begin by describing forensic work as set of tasks that are reasonably distinct. As a first attempt at such a classification, it would seem that what we refer under the broad ambit of "forensic science" would consist of the following tasks:

1. Basic Research. This would include activities such as developing new methods and technologies of forensic analysis, such as new chemical detection methods as well as the validation of these techniques. Some of this research occurs at universities or national laboratories,[106] and bench personnel, whose capacity is primarily dedicated to other tasks, conduct

---

105. *See generally* BETWEEN CRAFT AND SCIENCE, *supra* note 57; H.M. COLLINS, CHANGING ORDER: REPLICATION AND INDUCTION IN SCIENTIFIC PRACTICE (1985); MICHAEL POLANYI, THE TACIT DIMENSION (1966); STEVEN SHAPIN, A SOCIAL HISTORY OF TRUTH (1994); Stephen R. Barley & Beth A. Bechky, *In the Backrooms of Science: The Work of Technicians in Science Labs*, 21 WORK & OCCUPATIONS 85 (1994); Park Doing, *'Lab Hands' and the 'Scarlet O': Epistemic Politics and (Scientific) Labor*, 34 SOC. STUD. SCI. 299 (2004); Kathleen Jordan & Michael Lynch, *The Mainstreaming of a Molecular Biological Tool: A Case Study of a New Technique*, *in* TECHNOLOGY IN WORKING ORDER: STUDIES OF WORK, INTERACTION, AND TECHNOLOGY 162 (Graham Button ed., 1993).

106. *See generally* Victoria A. Smith et al., *The Reliability of Visually Comparing Small Frontal Sinuses*, 55 J. FORENSIC SCI. 1413, 1413-1680 (2010).

some of it.[107] The NAS Report is not critical of the quality of the current performance of this task, but it clearly states that there is not enough of it in terms of personnel, resources, or results.[108] It is fairly clear that U.S. funding agencies have tended to neglect forensic science.[109] Though basic researchers seem to have done reasonably well at validating new technologies and new detection methods, they have neglected one large area: the validation of pattern recognition techniques based on human visual interpretation (such as handwriting identification, bitemark identification, latent print identification, and firearms and toolmark identification, among other things). These activities are, of course, analogous to the sort of "discovery science" that has been the focus of most philosophical and early sociological accounts of "science."

2. *Evidence Collection*. This would include all activities that provide "inputs" to forensic laboratories, analyses, and assays. It is, of course, crucially important, in accord with the well-known "garbage in, garbage out" principle.[110] It is not clear how much scientific or philosophical knowledge is necessary to perform these tasks. Indeed, in many cases the task is performed by personnel without scientific training and without pretentions to being "scientists," such as sworn law enforcement officers.[111]

3. *Technical Management*. This would include work within a forensic laboratory overseeing the application of various forensic assays and ordering and coordinating various forensic assays in particular cases, but not necessarily actually performing the assays themselves. Increasingly, it

107. *See generally* Marcel de Puit, *An Alternative Trinity: Objectivity, Subjectivity, and Transparency*, 60 J. FORENSIC IDENTIFICATION 1 (2010).

108. *See* NAS REPORT, *supra* note 3, at 22.

109. *Id.*

110. A proverb from computing and information technology referring to the principle that nonsensical inputs will necessarily yield nonsensical output, attributed to the IBM computer technician George Fuechsel in the early days of computing. *See Garbage In, Garbage Out*, WIKIPEDIA, http://en.wikipedia.org/wiki/Garbage_in,_garbage_out (last visited Dec. 15, 2010).

111. KEITH INMAN & NORAH RUDIN, PRINCIPLES AND PRACTICE OF CRIMINALISTICS: THE PROFESSION OF FORENSIC SCIENCE 62 (2001) ("Perhaps the most bitter and persistent complaint throughout the history of criminalistics has been the lack of adequate training in crime scene procedures. In fact, most of the 'police science' or 'crime detection' books specifically address, at least in part, the audience of nonscientifically oriented police officers and detectives who are virtually always the first to arrive at a scene. More often than not, a law enforcement officer or evidence collection technician with minimal scientific training is the person tasked with the all important charge of recognizing and collecting evidence. Less and less often will a criminalist from the laboratory be called to the crime scene, and the decision to do so is usually that of those already there. The individual making decisions about what evidence to collect and the person given the responsibility to collect it vary widely between jurisdictions, so it is difficult to generalize." [citations omitted]).

might involve cost-benefit analyses of what assays should be conducted, taking into account not only the costs of the assays, but also the potential probative value of the results of such assays.[112] It might also involve piecing together the results of various assays in order to develop a holistic forensic "account" of the crime. Historically, technical management does not appear to have been a particularly prominent task and may not have even been explicitly defined as a task distinct from what I below call "Analysis." A number of converging trends, however, including increasing emphases on quality control, increasing attention to cost-benefit analyses, and heightened concerns about observer bias,[113] suggest that the profile of this task may be expected to grow.

4. Analysis. This would include all activities in which a forensic technique is deployed upon an item of evidence: such as a handwriting, bite mark, latent print, tool mark, DNA, or toxicological analysis.

5. Interpretation. This would include the interpretation and reporting of the meaning of an analysis or a set of analyses. As with "Technical Management," this task has not historically been regarded as particularly prominent or even as a task distinct from Analysis.[114] The NAS has, however, called for increasing attention to reporting of results, and recent scholarship has called attention to the crucial importance of this aspect of forensic work.[115] These trends suggest that we would do well to treat this as a distinct task.

Even based on this rudimentary definition of forensic tasks, it should be clear that the same principles, methods, and desired virtues are probably not appropriate to all five tasks. Furthermore, the standard principles, methods, and virtues that are habitually invoked for discovery science are not necessarily appropriate for all five tasks. It does not seem helpful, for example, to tell a scientific worker engaged in evidence collection to apply Popper's theory of falsificationism or to apply the "scientific method" as-

---

112. Christopher J. Lawless & Robin Williams, *Helping With Inquiries or Helping With Profits? The Trials and Tribulations of a Technology of Forensic Reasoning*, 40 SOC. STUD. SCI. 731 (2010).

113. Dan E. Krane et al., *Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation*, 53 J. FORENSIC SCI. 1006 (2008); D. Michael Risinger et al., *The* Daubert/Kumho *Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CAL. L. REV. 1 (2002).

114. NAS REPORT, *supra* note 3, at 111-26.

115. *See generally* Simon A. Cole, *Who Speaks for Science? A Response to the National Academy of Sciences Report on Forensic Science*, 9 LAW PROBABILITY & RISK 25 (2009); ; Dawn McQuiston-Surrett & Michael J. Saks, *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear*, 33 LAW & HUM. BEHAV. 436 (2009).

sociated with Sir Isaac Newton. Somewhat more controversially, I will also argue *infra* that it is not helpful to tell these things to a worker engaged in Analysis either. We may, however, want to impose a quality control regime upon workers engaged in Analysis, but we do not necessarily want to impose a quality control regime upon a basic researcher.

V. NORMATIVE GOALS FOR FORENSIC TASKS

If the standard set of principles, methods, and virtues invoked by the NAS Report are neither necessary nor helpful for all forensic tasks, then what principles, methods, and virtues *do* we need for each task? What "scientific culture" do we want for each task?

1. Basic Research. Presumably, we want forensic basic researchers to behave like basic researchers in other areas of science. We want them to innovate and to subject their innovations to rigorous scrutiny. We want the individuals who do this work to be trained scientists, much as they are in other areas of science. We want a "culture" that is as much like a university—or at least like an industrial research laboratory—as possible.

2. Evidence collection. Here, we do not want a culture centered around hypothesis testing, falsificationism, or organized skepticism. What seem to be most needed are care, accountability, meticulous documentation, and ethics. By far the greatest concern expressed in the forensic literature about evidence collection concerns the potential for contaminating the crime scene.[116] It seems that the main concern is that we want people who are careful, meticulous, and honest. They need not know philosophy of science or be competent to practice science. Primarily, they need to be aware of the current capabilities and limitations of forensic science, so as to know what to collect and what not to collect, what sorts of actions might lead to contamination, and so on. In this area, therefore, we would expect "scientific culture" to mean something very different: care, meticulousness, and attention to detail, among others.

3. Technical management. Here we want an individual to think scientifically about a case as a whole. Due to the widespread popularity of Popper, many have suggested that this "thinking" should be conceptualized in terms of hypothesis testing.[117] This may not necessarily be wrong, but it is probably not ideal. Contemporary philosophers of science would probably suggest that the Technical Manager should evaluate which explanation of the case appears most robust as evidence emerges. Among scholars who think about forensic science, however, the dominant approach has been to

116. INMAN & RUDIN, *supra* note 111.
117. *See supra* note 87.

evaluate a case from a Bayesian framework.[118] Thus, a Technical Manager would need to be aware of the capabilities and limitations of the tools used by Analysts (whether or not she has practiced as an Analyst herself) and would need to know some philosophy of science, most likely a Bayesian approach given the current state of the field. The Technical Manager would not, however, need to know how to do basic research. The "scientific culture" would be one of rigorous open-mindedness and critical thinking with an emphasis on avoiding traps in reasoning, such as circularity and the transposed conditional.[119]

4. *Analysis*. While there have been attempts to conceptualize forensic analyses in terms of hypothesis testing, as discussed above,[120] this is neither necessary nor, probably, appropriate. Instead, we should think of analysts much in the way we think about laboratory technicians in a university, industrial, or medical setting. We do not want, or need, these analysts to think about the validity of the assays they perform. What we want is for analysts to do well at performing assays that someone else has validated.[121] We want them to be careful, meticulous, and honest in their application of these assays. We want them to document their work, and to adhere to protocols. The "scientific culture" we want has far more to do with care, meticulousness, documentation, and honesty than with hypothesis testing, falsification, or organized skepticism.

5. *Interpretation*. As numerous scholars have documented, there are a host of tricky issues raised by the interpretation of forensic evidence.[122] Handling these issues requires expertise in logical reasoning that seems to most commonly be acquired through training in philosophy, law, science, mathematics, or statistics. Here "scientific culture" would essentially be centered around logical reasoning of the highest order. The primary values might be a determination to report the evidence as accurately and precisely as possible, as well as a sense of self-restraint that would allow the Inter-

---

118. *See, e.g.*, COLIN G.G. AITKEN, STATISTICS AND THE EVALUATION OF EVIDENCE FOR FORENSIC SCIENTISTS (1995); BERNARD ROBERTSON & G. A. VIGNAUX, INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM (1995).

119. *See generally* William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials*, 11 LAW & HUM. BEHAV. 167 (1987) (explaining fallacies in the interpretation of statistical evidence to which lay people are susceptible).

120. *See supra* note 87.

121. This is a broad generalization that would need to be discussed more specifically for different assay, laboratories, countries, and so on. For some assays, "local" validation of specific instruments is necessary. This is a separate issue from what we might think of as "global" validation of the ability of certain technique to detect certain things.

122. *See generally* AITKEN, *supra* note 118; DAVID J. BALDING, WEIGHT-OF-EVIDENCE FOR FORENSIC DNA PROFILES (2005); ROBERTSON & VIGNAUX, *supra* note 118.

preter to resist the temptation to make more inferences than are warranted by the analytic results. Interpreters would have to think much more carefully about probability and proof than do Basic Researchers, and they would have little use for the skills and virtues associated with Basic Researchers.

VI. THE CURRENT STATE OF AFFAIRS

Having laid out an ideal model for forensic science, let us see how the current state of affairs, as the NAS Report found it, measures up. In most of the pattern recognition disciplines, the various task categories articulated above are blurred, blended, and often performed by the same individual. Some basic research is performed in university, industrial, or government laboratories, but not enough. Therefore, it has been left to a few intellectually curious analysts to perform the balance of basic researchers, often in their personal time since they are still expected to perform their normal workload of analysis.[123] It is not to demean the contributions of these self-sacrificing individuals to say that they are often not trained to do basic research, lack the resources typically available to researchers at university, industrial, and government laboratories, and lack the professional networks basic researchers use to test their research and generate innovation.[124]

Evidence collection, in contrast, *is* sometimes separated as a task from the other aspects of forensic science.[125] But, in many other cases, the same individual performs evidence collection and analysis.[126]

The task I have called here "technical management" is often not performed at all and is sometimes handled by a non-scientist detective or an experienced analyst who has become a manager.

---

123. Examples of this in pattern recognition include Tuthill and Ashbaugh, practitioners with virtually no formal scientific education who essentially remade the conceptual foundations of pattern recognition. *See generally* ASHBAUGH, *supra* note 51; HAROLD TUTHILL, INDIVIDUALIZATION: PRINCIPLES AND PROCEDURES IN CRIMINALISTICS (2004). The achievements of these individuals are all the more remarkable given their *lack* of formal scientific education; my point is not to criticize their impressive achievement, but rather to argue that it is a poor institutional structure that relies on such individuals for basic research.

124. *Id.*; NAS REPORT, *supra* note 3, at 14 ("[F]orensic science and forensic pathology research, education, and training lack strong ties to our research universities and national science assets.").

125. INMAN & RUDIN, *supra* note 111.

126. ASHBAUGH, *supra* note 51, at 136; Pat A. Wertheim, *Integrity Assurance: Policies and Procedures to Prevent Fabrication of Latent Print Evidence* 48 J. FORENSIC IDENTIFICATION 431, 435 (1998); Interpol European Expert Group on Fingerprint Identification, *Method for Fingerprint Identification*, INTERPOL (2005), *available at* http://www.interpol.com/public/Forensic/fingerprints/WorkingParties/IEEGFI/ieegfi.asp# (last visited Dec. 2, 2010).

Analysis and interpretation are currently not separated at all for pattern recognition disciplines. The notion that these tasks should be separated is simply a proposal; it does not currently exist.[127]

And then there is the crucial issue of validation. The NAS Report noted that validation was lacking for nearly all of the non-DNA forensic identification technologies.[128]

A. Historical Explanation for the Current State of Affairs

How did this happen? How did we come to a point where techniques were used in court, in some cases for a century, without any validation studies?[129] How were Analysts permitted to report conclusions in a way that, the NAS Report now tells us, was unsupported for decades?[130] It would seem that we can explain the current state of affairs historically—that we can discern historically how we arrived at the position we are in today.

Basic researchers, who largely set their own research agendas, never set out to do validation studies. There were probably a number of reasons for this. There was little intellectual incentive to conduct validation studies of techniques that were already being used and had been accepted as valid by courts.[131] Funding agencies provided little material incentive to do such research.[132] Most forensic basic researchers tend to have expertise in a

---

127. This is actually somewhat curious. For example, one of the oft-noted problems with latent print analysis is that there is no metric for "sufficiency"; no objective measurement that would allow the Interpreter to know that the amount of friction ridge detail found consistent between two prints is sufficient to conclude that one should expect only one piece of friction ridge skin in the universe to produce prints consistent with that detail. Instead, the practice is to rely on Analysts to make intuitive judgments about the rarity of various configurations of friction ridge detail in the universe. This is obviously inferior to estimates of rarity based on objective data from sampled populations. Nonetheless, the system might be improved by separating analysis from interpretation—that is, having one experienced latent print analyst determine whether the ridge detail is consistent between two prints, and then asking another experienced analyst to make a separate intuitive estimate of the rarity of the consistent detail.

128. NAS REPORT, *supra* note 3, at 87.

129. *Id.* at 43.

130. *Id.* at 87.

131. Simon A. Cole, *Is Fingerprint Identification Valid? Rhetorics of Reliability in Fingerprint Proponents' Discourse*, 28 LAW & POL'Y 109, 129 (2006).

132. There is the now well-known scandal of the 2000 National Institute of Justice Request for Proposal ("RFP") for fingerprint validation research. The RFP was allegedly suppressed by the FBI in order to avoid jeopardizing its position in an admissibility hearing on fingerprint evidence and the proposal was never funded. By all accounts, one of the Principal Investigators on the most credible proposal was retained by the defendant in that same admissibility hearing. NAT'L ACADEMIES, REPORT TO CONGRESS (2009), http://www.nationalacademies.org/annualreport/eng09.html (last visited Dec. 2, 2010).

specific scientific discipline, such as chemistry or biology,[133] which may be appropriate for developing new detection methods, but not necessarily for conducting validation studies. As FBI Laboratory Director Christian Hassell put it, validation research is "the 'valley of death' because 'nobody wants to pay for it, nobody really wants to do it.'"[134]

Because the mainstream scientific community historically showed little interest in forensic science, the "validation question" was never posed within the scientific community. It was left to criminal defendants to ask, "Where is the study showing this technique is valid?" The *Daubert* decision in 1993[135] provided an opening for defendants to ask precisely this question.

Because analyses have been blurred with interpretation, it was the Analysts who appeared in court. And, so the question, "Where is the study showing this technique is valid?" was posed to Analysts, not to Interpreters (who barely, if at all, exist),[136] Technical Managers, or Basic Researchers. Basic Researchers had long been focused on such matters as developing new detection techniques and therefore had not built up either a body of knowledge, or even thinking, about the question of validation.[137] Moreover, connections between Basic Researchers and Analysts were weak.[138] For all these reasons, Basic Researchers were of little use as resources when the validation question was posed to Analysts in court. Analysts were left to answer the question alone.

An intellectually valid answer to the question "Where is the study showing this technique is valid?" requires some familiarity with scientific reasoning and probably some understanding of philosophy of science as well. Historically, Analysts had been drawn either from the ranks of police or civilian law enforcement employees.[139] They did not have the kinds of scientific training that would allow one to function as a Basic Researcher in a university, industrial, or government laboratory, and some had no scientific training at all.

---

133. Dep't of Chemistry, Cal. State Univ., Fresno, *Forensic Science*, http://www.csufresno.edu/chem/facstaff/forensic.shtml (last visited Dec. 2, 2010).

134. Laura Spinney, *Forensic Science Braces for Change*, NATURENEWS (July 22, 2010), http://www.nature.com/news/2010/100722/full/news.2010.369.html.

135. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993).

136. One might argue that forensic consultants, who reinterpret the results of other Analysts, are Interpreters.

137. *See* Nat'l Academies, *supra* note 19.

138. NAS REPORT, *supra* note 3, at 14.

139. David L. Grieve, *The Identification Process: Traditions in Training*, 40 J. FORENSIC IDENTIFICATION 195 (1990).

Under these circumstances, Analysts offered answers in court to the validation question that would not have been either offered or accepted by a professional scientist. These answers included the following: what I have called the "fingerprint examiner's fallacy," that the claimed uniqueness of the target of analysis demonstrates the validity of the analysis;[140] what Saks and Koehler have called the "individualization fallacy," that a finding of consistency between two images warrants the conclusion that the images must derive from the same source;[141] what I have called "casework validation," that longevity of use in court constitutes proof of the accuracy of the technique;[142] and the separation of error rates into "methodological" and "human" categories such that the "methodological error rate" can be meaningfully designated as zero.[143]

These answers were unfortunate but predictable outcomes of the social structure of forensic science in the pattern recognition disciplines. The courts got as good answers as they could have expected from a group of Analysts with the training, background, orientation, and, yes, *culture* that they had. But what really compounded the problem was that the courts accepted these answers hook, line, and sinker.[144] There ensued then a decade long debate in the pages of law journals and legal opinions.[145] Nearly a decade was lost seeking to establish that the uniqueness of friction ridge skin was not the right empirical question to ask about the validity of latent print identification, that use in court could not substitute for scientific validation testing, and that it was misleading to call the error rate of forensic assays "zero." While these canards would seem to have been put to rest by the NAS Report—which rejected them all—in fact, so firmly are they now ensconced in Analysts' and courts' way of thinking that they are proving very difficult to dislodge, and they continue to live on.[146]

My point is less that Analysts gave poor answers than that Analysts should never have been asked to defend the epistemological underpinnings

---

140. Simon A. Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 LAW PROBABILITY & RISK 233, 236 (2009) [hereinafter Cole, *Forensics Without Uniqueness*].

141. Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199, 204-05 (2008).

142. Simon A. Cole, *'Implicit Testing': Can Casework Validate Forensic Techniques?*, 46 JURIMETRICS 117 (2006).

143. Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985 (2005); *see also* NAS REPORT, *supra* note 3, at 143.

144. MODERN SCIENTIFIC EVIDENCE, *supra* note 13.

145. *Id*.

146. Cole, *Forensics without Uniqueness*, *supra* note 140.

of their assays. Given the social structure of forensic science, there was no alternative; there was no one else. That a mess resulted should surprise no one. The situation was somewhat analogous to haling radiological technicians into court to testify about the epistemological validity of radiological examinations. Of course, that would never happen because radiological technicians could draw on the support of physicians who interpret the results in light of medical knowledge and the researchers who developed and validated the technologies.

Understood in this light, we can see that the failure of scientific culture that the NAS Report implies cannot be understood as a unitary failure that applies equally to all task-roles in forensic science. Nor is it a failure to adhere to a single method, principle, or virtue. Instead, the implied failure of scientific culture should be understood as a much more variegated thing. For instance, the failure to validate most of the pattern recognition forensic identification techniques can be understood as a cultural failure of intellectual curiosity among Basic Researchers. The use of non-validated technologies can be understood as a cultural failure to engage in organized skepticism, or, in the NAS Report's words, "critical questioning."[147] The creation of reporting regimes which mandated the reporting of conclusions in terms that were illogical and unsupported (e.g., "individualization") can be understood as a cultural failure of epistemological modesty, a failure of, in the NAS Report's words, a "scientific culture [that] encourages cautious, precise statements and discourages statements that go beyond established facts."[148] And, the defensiveness with which the forensic community reacted to the challenges posed by outsiders—challenges that, according to the NAS Committee, turn out to have been, in fact, warranted—can be understood as a cultural failure of "openness to new ideas, including criticism and refutation," not to mention a failure of the virtue of collegiality.[149]

VII. BUILDING A FORENSIC SCIENTIFIC CULTURE

It should now be clear that building a scientific culture in forensic science will require more than a blanket recitation of Popperian theories and Mertonian virtues. Instead, we need to think carefully about the roles played by various actors in forensic science, the virtues desired by each, and the principles and methods we want each to follow. We need not, however, necessarily start this process of thinking from scratch. Other areas of science—arguably all areas—accommodate multiple roles. We

147. NAS REPORT, *supra* note 3, at 125.
148. *Id.*
149. *Id.* at 113.

can perhaps draw on other areas of science for analogies that might help suggest models for an ideal social structure of forensic science.

Medicine would seem to pose a model of obvious relevance to forensic science.[150] Medicine, as a broad category, encompasses a broad range of technoscientific workers who play a variety of roles. For our purposes, however, we can focus on four roles that are analogous to the roles we defined for forensic science. First, there are biomedical researchers, who tend to hold Ph.D.'s, M.D.'s, or both. These researchers are engaged in basic research and the production of knowledge about the natural world. They may never see patients, never have seen a patient since medical school (like Oliver Sacks in a memorable scene in *Awakenings*), or (in the case of Ph.D.'s) never have seen a patient at all.[151] They may not be competent to administer medical treatments or perform medical procedures.[152]

Second, there are clinical physicians. Crucially, these individuals must be competent to read, digest, and apply the medical knowledge that is produced by medical researchers, as it is disseminated to them through such mechanisms as journals, conference presentations, (less optimally) pharmaceutical company advertisements, and so on.[153] These individuals are ethically and legally required to make decisions in light of the current knowledge produced by researchers.[154] They need not, however, perform

150. I am hardly the first to suggest this. Dr. Stoney anticipated some of the arguments made here, though his focus is on forensic science education, rather than its institutional structure. David A. Stoney, *A Medical Model for Criminalistics Education*, 33 J. FORENSIC SCI. 1086 (1988). Another forthcoming article, on which I am a co-author, also explores themes quite similar to those mentioned here. Jennifer L. Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, UCLA L. REV. (forthcoming 2011). Not surprisingly, there may be some differences in nuance between that, jointly authored article and this sole-authored one, but the general thrust of the arguments is largely consistent.

151. OLIVER SACKS, AWAKENINGS (1973); Molly Cooke et al., *American Medical Education 100 Years After the Flexner Report*, 355 NEW ENG. J. MED. 1339 (2006).

152. Cooke et al., *supra* note 151.

153. HARRY H. MARKS, THE PROGRESS OF EXPERIMENT: SCIENCE AND THERAPEUTIC REFORM IN THE UNITED STATES, 1900-1990, at 231 (1997) ("Physicians are presumed to accept voluntarily either the rational dictates of scientific method or the judgments of constituted authorities."); Eliot Freidson, *The Reorganization of the Medical Profession*, 42 MED. CARE RES. REV. 11, 30-31 (1985) [hereinafter Freidson, *Reorganization*] ("Where once all practitioners could employ their own clinical judgment to decide how to handle their individual cases independently of whatever medical school professors asserted in textbooks and researchers in journal articles, now the professors and scientists who have no firsthand knowledge of those individual cases establish guidelines that administrators who also lack such firsthand experience attempt to enforce.").

154. MARKS, *supra* note 153, at 231 ("[M]embership in the republic of science was offered to those who would acknowledge the constituted authorities within medicine by allowing their deliberations and reflections to serve as a surrogate for the judgments of the individual physician."); Eliot Freidson, *The Changing Nature of Professional Control*, 10 ANN.

research themselves or even be competent to perform research.[155] Therefore, often they will "know" that a treatment has a given degree of effectiveness under given conditions because of some research.[156] What this means is that research findings have been disseminated to them in some form which states that some researcher has determined the treatment has that degree of effectiveness under those conditions. The clinician does not, of course, "know" the effectiveness of the treatment in the sense of having performed the research herself or even having seen it done. What we have in place in medicine is a trust-based system of knowledge dissemination, with trust invested in scientific institutions like journals and their peer review systems. While such systems are notoriously far from foolproof,[157] they have been adopted as reasonable pragmatic solutions to the problem of knowledge dissemination in socially important areas like medicine.

Third, there are laboratory technicians who perform a variety of procedures and tests, from biochemical assays to radiological imaging. These individuals rarely, if ever, hold M.D.'s or Ph.D.'s.[158] Although they do not have the knowledge that clinicians have acquired by attending medical school and functioning as physicians, a long tradition of sociological research has shown that these individuals often have very sophisticated knowledge that takes other forms.[159] In some cases, this knowledge has been referred to as "tacit knowledge."[160] Some laboratory technicians may be more competent than physicians at performing certain laboratory procedures, and some radiological technicians may be more skilled at reading

REV. SOC. 1, 1 (1984) [hereinafter Freidson, *Changing Nature*] ("Rank and file practitioners are no longer as free to follow the dictates of their individual judgments as in the past . . . .").

155. Freidson, *Changing Nature*, *supra* note 154 ("The technical standards" governing physicians "are devised by a separate group of professionals—the knowledge elite—who are based primarily in professional schools . . . .").

156. Stephen Harrison & Waqar I.U. Ahmad, *Medical Autonomy and the UK State 1975 to 2025*, 34 SOCIOLOGY 129, 138 (2000) ("[I]t has increasingly become the case . . . that clinical decisions must be justified by reference to external research findings . . . .").

157. *See generally* FIONA GODLEE & TOM JEFFERSON, PEER REVIEW IN HEALTH SCIENCES (2003).

158. *See generally* Stephen R. Barley & Beth A. Bechky, *In the Backrooms of Science: The Work of Technicians in Science Labs*, 21 WORK & OCCUPATIONS 85 (1994); Stephen R. Barley, *Technology as an Occasion for Structuring: Evidence From Observations of CT Scanners and the Social Order of Radiology Departments*, 31 ADMIN. SCI. Q. 78 (1986); Stephen R. Barley, *The Social Construction of a Machine: Ritual, Superstition, Magical Thinking and Other Pragmatic Responses to Running a CT Scanner*, *in* BIOMEDICINE EXAMINED 497 (Margaret Lock & Deborah R. Gordon eds., 1988).

159. POLANYI, *supra* note 105.

160. *Id*.

images than their clinician supervisors.[161] Crucially, however, even when technicians possess superior skills, we do not generally expect them to make diagnoses. The task of diagnosis is reserved for clinicians, largely because of the knowledge that clinicians acquire from researchers. Even if a radiological technician is more skilled at interpreting a film that the physician, we believe the physician is able to place that interpretation within the context of medical knowledge. In this sense, we want medical technicians to exercise their manual and interpretive skills, but we do not want them to make inferences about what those interpretations mean for the larger "case" as a whole, the diagnosis of the patient. Moreover, technicians' knowledge about the validity of the techniques they apply is typically very limited.[162] Medical technicians would probably be quite hard pressed to cite the studies which validated the tests or instruments that they use.

This division of medical labor was not always in place, but is the product of historical changes in medical education and practice. In the late nineteenth century, medical education famously became more "scientific."[163] These educational changes were based on the presumption that "[a]lthough 'not every student . . . can become an experimenter . . . every physician must be so educated that he may intelligently apply the knowledge furnished him by experimental medicine in the cure of such diseases as can be cured."[164] This is precisely the principle I am advocating for the relationship between Researchers and Technical Managers.

At this point, the thrust of the analogy should be clear: in forensic science, we would want a cadre of basic researchers, who develop and validate new methods and techniques. We would also want a much larger cadre of technicians with manual and interpretive skills. These individuals would need to know very little about the validity of the techniques that they use. What we primarily want from them would be to exercise their skills well. Mediating between these groups would be a cadre of individuals with more scientific training and knowledge than technicians. These individuals would need to know enough science to be educated consumers of the knowledge produced by basic researchers. They would, however, not necessarily need the set of skills necessary to be independent basic researchers themselves. These individuals would presumably function as laboratory technical managers. They would know whether certain techniques are va-

161. *See supra* note 158.

162. *See supra* note 158.

163. THOMAS NEVILLE BONNER, BECOMING A PHYSICIAN: MEDICAL EDUCATION IN BRITAIN, FRANCE, GERMANY, AND THE UNITED STATES, 1750-1945 (1995).

164. *Id.* at 290-91; *see also* MARKS, *supra* note 153.

lid or not, not because they had validated them themselves or even seen it done, but because they would understand what a validation consists of, and would be capable of evaluating scientific literature to determine whether the studies that had been conducted were appropriate to support the claimed results.

The fourth role in forensic science is filled by crime scene technicians. I have not focused on this group because I don't think it is the focus of the issues identified by the NAS Report. For purposes of ensuring the neatness of the analogy with medicine, however, we can compare crime scene technicians to Emergency Medical Technicians ("EMTs"). Although the analogy is not perfectly apt, there are some similarities: they operate in the field; they do not have the broad base of medical knowledge held by physicians (or even nurses), but do deploy a small, highly specific corpus of medical knowledge; they operate according to strict ethical and legal controls regarding what they can and cannot do at "the scene."[165]

A. Hierarchy

Medicine is notoriously hierarchical, and, in drawing an analogy with medicine, we should be candid about the fact that we are proposing a hierarchical structure for forensic science. Hospitals are, of course, notoriously hierarchical, with physicians at the top of the hierarchy. The relationship between physicians and medical technicians, for example, is obviously hierarchical. There is, however, a sense in which the relationship between medical researchers and clinicians is hierarchical as well. I do not mean this in a political sense, so much as an epistemological sense. Clinicians are, in some way, required to consume the knowledge that medical researchers produce. The nature of this "requirement" has, of course, changed greatly in the last couple of decades and remains greatly in flux as evidence-based medicine ("EBM") and clinical practice guidelines ("CPGs") have become more common in medicine.[166] But even prior to the development of EBM there was an expectation that clinicians were required to know what was in the literature.[167] As they do today, clinicians retained a great deal of autonomy and discretion in applying what was in the literature,[168] yet there were points at which a treatment became so dis-

165. STEFAN TIMMERMANS, SUDDEN DEATH AND THE MYTH OF CPR 184-205 (1999).

166. *See generally* STEFAN TIMMERMANS & MARC BERG, THE GOLD STANDARD: THE CHALLENGE OF EVIDENCE-BASED MEDICINE AND STANDARDIZATION IN HEALTH CARE (2003).

167. Freidson, *Reorganization*, *supra* note 153.

168. Freidson, *Changing Nature*, *supra* note 154.

credited in the literature that it would be malpractice to apply it (or, vice versa, so accepted that it would be malpractice not to apply it).[169]

Of course, no one suggests that this has been a recipe for harmony in medicine. To the contrary:

> Since the standards of the knowledge elite are grounded in the abstract world of logic, scientific principles, and statistical probabilities rather than in the concrete world of work, in experimental designs and controlled laboratory findings rather than in the untidy, uncontrolled arena of practice, and in circumstances that are considerably less subject to the constraints of time, money, equipment, and other resources than is true of everyday practice, it is not hard to understand the skepticism of the rank and file professional.[170]

So we should be candid: the medical model we are proposing here is not a recipe for harmonious convergence. Rather, it is my assertion that there is a set of limited options: the status quo in which non-scientifically-educated practitioners are left to fend for themselves, a "harmony" model in which non-scientifically-educated practitioners and scientifically-educated researchers (and perhaps even lawyers and police) are treated as equal "stakeholders," and a "hierarchical" model in which a "knowledge elite" of researchers exerts control over practitioners. We suggest being candid that, while hierarchy may not always be the most palatable thing to advocate, it appears to be the best model for society to get what it wants from forensic science, and it is the least bad option. It should be noted, for example, that society is reasonably content with the hierarchical model in medicine. While there is certainly great public resentment when a rank and file physician's discretion is limited by a non-physicians, such as an insurer, accountant, or medical administrator, there is far less public sentiment behind the notion that primary care providers should be permitted to ignore the findings of medical research.[171]

We propose here the same sort of hierarchy for forensic science. Researchers would have the last word on whether a method or technique is valid. Technicians would no longer be put in the awkward position of having to defend the validity of the techniques they apply. Likewise, they would no longer have the power, as they have had for so long, to *ipse dixit* "dec-

169. *Id.* at 2.

170. *Id.* at 16.

171. *See, e.g.*, Physicians Warned Not to Ignore Suicide Clues, National Center for Biotechnology, Newsbriefs, 132 CAN. MED. ASSOC. J. 1404 (1985), *available at* http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1346111/pdf/canmedaj00263-0070.pdf.

lare" or vouch for[172] the validity of the techniques they apply. Technical Managers, meanwhile, might be called upon to explain the validity of the techniques used in their laboratories, but they would do so based on research conducted by Basic Researchers.[173] Technical Managers, however, would be *required* to be cognizant of, and take into account, the scientific literature, much in the way that medical clinicians are. Technical Managers would, in a sense, be *required* to consume basic forensic scientific knowledge.

In proposing hierarchy, it should be noted that we are not proposing the creation of an elite "priesthood" that would have a monopoly on the legitimation of knowledge.[174] Recall that this vision would assume, as a fundamental precondition, the removal of NIFS and of forensic laboratories from the control of law enforcement. Basic Researchers would be expected to be a diverse group of scientists with diverse viewpoints, as medical researchers are today. While some of them might have "official" positions in forensic science (e.g., NIFS scientists), others might be independent of forensic science and employed by universities, industrial corporations, or non-government organizations. There would be an interaction between government scientists and those outside of government, much as there is between scientists who work for governmental scientific institutions like the National Institutes of Health or the Food and Drug Administration and

172. *See generally* Michael J. Saks, *Banishing* Ipse Dixit*: The Impact of* Kumho Tire *on Forensic Identification Science*, 57 WASH. & LEE L. REV. 879 (2000).

173. We might distinguish here between the validation of instruments and the validation of what I call assays. All forensic assays require general validation. A study is necessary to show that, for example, a mass spectrometer is capable of distinguishing one particular set of chemicals from others. Many assays also require what we might call specific validation. In forensic disciplines involving instrumental analysis, such as DNA profiling and drug analysis, it is often necessary to validate instruments locally as well. For example, having established that mass spectrometers in general are capable of detecting particular sets of chemicals, laboratories typically also need to establish that a particular mass spectrometer is capable of detecting these chemicals, as used in the local laboratory by the local laboratory personnel. These specific validation activities would still need to be performed by Technical Managers or perhaps Analysts. I am, however, suggesting that general validation activities should not be located in a front-line laboratory.

The situation is somewhat different in the pattern recognition disciplines, such as latent prints, handwriting analysis, bite mark analysis, and so on, in which the "instrument" is a trained human being. In these disciplines, there is no specific validation of the "instrument," and thus there would be no need for Analysts to engage in validation activities. Of course, one might think of quality control mechanisms as the equivalent of specific validation. Such activities would presumably be performed (or at least supervised) by Technical Managers and Laboratory Directors.

174. Roger G. Koppl et al., *Epistemics for Forensics*, 5 EPISTEME 141, 154 (2008).

those who work outside of government.[175] While such relationships are naturally fraught, problematic, and conflicted, with potential dangers that conventional wisdoms get entrenched or corporate interests get served,[176] I argue that they are less problematic than the status quo in forensic science.

B. The Deskilling of Forensic Science

We also need to be candid about the fact that what we are proposing represents, in some sense a "deskilling"[177] of the profession of forensic science.[178] We are proposing to break the task of the forensic scientist, as classically understood, into segments that would be assigned to different individuals with different skill sets, educational backgrounds, expectations, and roles. We are proposing that some of the individuals, particularly the technicians, simply will not need to know, or even think much about certain things, and this may be construed as countenancing ignorance.

We need to be candid about this because an alternative, and reasonable, normative proposal exists which would move things in precisely the opposite direction. This is a view that I associate most closely with generalist forensic scientists with research focused scientific educations like Inman, Rudin, and DeForest. This view is also loosely associated with the University of California, Berkeley forensic science program run by Professor Paul Kirk. For convenience, we can refer to them as the "California School."[179] The California School might argue that, rather than deskilling and segmenting the profession, we should be uplifting it. While they might agree that it was a mistake to expect technicians without significant scientific training to defend, or even talk or think coherently about the validation of techniques like latent print or firearms and toolmark identification, they might argue that the answer is not to keep those non-scientifically trained individuals in the technician role. Rather, they might argue that the goal should be to turn all persons occupying the role of "forensic scientist" into true scientists with a scientific approach to empirical questions—in short, a "scientific culture." This is radically different proposal than the one I outlined above, in which, rather than differentiating roles, essentially, everyone in forensic

175. *See, e.g.*, ARTHUR A. DAEMMRICH, REGULATORY LAWS AND POLITICAL CULTURE IN THE UNITED STATES AND GERMANY, *in* REGULATION OF THE PHARMACEUTICAL INDUSTRY (JOHN ABRAHAM & HELEN LAWTON SMITH eds., 2003).

176. *Id.*

177. *See* HARRY BRAVERMAN, LABOR AND MONOPOLOGY CAPITAL: THE DEGRADATION OF WORK IN THE TWENTIETH CENTURY 424-47 (1974).

178. INMAN & RUDIN, *supra* note 111.

179. This is not to say that Inman, Rudin, and DeForest have identical views, but rather somewhat similar views that might be traced to a common orientation and approach.

science would have the skills and knowledge of at least Technical Managers and perhaps Basic Researchers. These Technical Managers would then approach *each case* as a scientific problem to be solved and deploy available forensic scientific tools as necessary to enhance our understanding of what might have occurred at the scene of the crime. Perhaps the best way to illustrate the difference between this mindset and the one that guides the proposal I outlined above, is to refer to DeForest's defense of "[o]ne person labs."[180] DeForest's claim, made not that long ago, that having a single individual with scientific training is a reasonable structure for forensic laboratories in small jurisdictions illustrates a very different conception of forensic science. In this view, cases are approached as problems by a single individual who possesses all of the tools encompassed by the term "forensic science"—bench skills, conceptual tools, and research background—and deploys these skills as appropriate to solve the problem.

This too is a compelling vision. I want to emphasize that I am agnostic as to whether it would be better to segment and differentiate the profession, as I have proposed above, or to uplift the entire profession to the doctoral (or near-doctoral) level. I think that having every individual working in forensic science trained to the level of Inman, Rudin, and DeForest would, by itself, solve most of the "the problems that plague the forensic science community."[181] I think that, to pose a historical counterfactual, had the profession developed so that everyone in it had the scientific training of Inman, Rudin, and DeForest, no one would ever have claimed that the error rate of latent print identification was "zero," and we would not have an NAS Report today.[182]

To be clear, I think either solution—differentiation or uplift—would be acceptable. I must note, however, that it seems clear to me that differentiation is a much more realistic proposal because it is more in tune with trends already widespread in forensic science that are independent of the issues raised by the NAS Report. These trends would include cost-cutting (the "uplift" strategy would be extremely expensive), managerial efficiency,

---

180. Peter R. DeForest, *Recapturing the Essence of Criminalistics*, 39 SCI. & JUST. 196, 198 (1999).

181. Edwards, *supra* note 23.

182. I want to distinguish here between a philosophical principle, and a general impressionistic statement about the world. I am not claiming that the possession of a doctorate necessarily and automatically would prevent someone from making silly assertions, like "the error rate of latent print identification is zero" or "latent print identification is valid because it has been used in court for one hundred years." I am merely making the rhetorical point that I believe, based on little more than my own "sense of things," that, had the profession of forensic science been entirely staffed by individuals with doctorate, such uncharacteristic statements would have been far less likely to have appeared.

consolidation of laboratories, and so on. So, while I would applaud a mandate from Congress that all forensic science be performed by doctoral-level forensic scientists, I do not think such a proposal is realistic.

As sociologist of medicine Eliot Freidson notes, it is precisely this hierarchy of control through *internal differentiation* within the profession (i.e., a "knowledge elite" and rank and file practitioners) that makes medicine a "profession" rather than a "craft." Crafts, in contrast, are subject to *external* regulation and control: "Without physicians serving in both roles, the profession could only sustain a position that is at best like that of the crafts, dependent on its organization but at the mercy of others' technological innovations and administrative practices."[183] Thus, I would suggest, it is perhaps only through adoption of this medical model that forensic science can live up to the subtitle of Inman and Rudin's book, "The Profession of Forensic Science."[184]

183. Freidson, *Reorganization*, *supra* note 153, at 30.
184. INMAN & RUDIN, *supra* note 111.

Jennifer L. Mnookin, Simon A. Cole, Itiel E. Dror, Barry A. J. Fisher, Max M. Houck, Keith Inman, David H. Kaye, Jonathan J. Koehler, Glenn Langenburg, D. Michael Risinger, Norah Rudin, Jay Siegel, and David A. Stoney, "The Need for a Research Culture in the Forensic Sciences" 58 UCLA L. R. 725 (2011).
Reproduced with permission.

THE NEED FOR A RESEARCH CULTURE IN THE FORENSIC SCIENCES

Jennifer L. Mnookin, Simon A. Cole, Itiel E. Dror, Barry A. J. Fisher, Max M. Houck, Keith Inman, David H. Kaye, Jonathan J. Koehler, Glenn Langenburg, D. Michael Risinger, Norah Rudin, Jay Siegel, and David A. Stoney*

The methods, techniques, and reliability of the forensic sciences in general, and the pattern identification disciplines in particular, have faced significant scrutiny in recent years. Critics have attacked the scientific basis for the assumptions and claims made by forensic scientists both in and out of the courtroom. Defenders have emphasized courts' longstanding acceptance of forensic science evidence, the relative dearth of known errors, and practitioners' skill and experience. This Article reflects an effort made by a diverse group of participants in these debates, including law professors, academics from several disciplines, and practicing forensic scientists, to find and explore common ground. To what extent do the forensic sciences need to change in order to place themselves on an appropriately secure foundation in the twenty-first century? We all firmly agree that the traditional forensic sciences in general, and the pattern identification disciplines, such as fingerprint, firearm, toolmark, and

* Jennifer L. Mnookin, Professor of Law, Founding Co-Director of the Program on Law, Science, and Evidence (PULSE), UCLA School of Law; Simon A. Cole, Associate Professor, Department of Criminology, Law & Society, UC Irvine; Itiel E. Dror, Researcher, Institute of Cognitive Neuroscience, University College, London, and Principal Consultant, Cognitive Consultants International (CCI); Barry A. J. Fisher, Crime Laboratory Director (retired), Los Angeles County Sheriff's Department; Max M. Houck, Director, Forensic Science Initiative, Research Office, and Director, Forensic Business Research and Development, College of Business and Economics, West Virginia University; Keith Inman, Assistant Professor of Criminal Justice Administration, Cal State University, East Bay, and Senior Criminalist, Forensic Analytical Sciences, Inc.; David H. Kaye, Distinguished Professor and Weiss Family Scholar, Dickinson School of Law, and Graduate Faculty, Forensic Science Program, Eberly College of Science, Penn State University; Jonathan J. Koehler, Beatrice Kuhn Professor of Law, Northwestern University School of Law; Glenn Langenburg, Forensic Scientist, Minnesota Bureau of Criminal Apprehension Forensic Science Services; D. Michael Risinger, John J. Gibbons Professor of Law, Seton Hall University School of Law; Norah Rudin, Forensic DNA Consultant; Jay Siegel, Chair, Department of Chemistry and Chemical Biology and Director, Forensic and Investigative Sciences Program, Indiana University–Purdue University; David A. Stoney, Chief Scientist, Stoney Forensic, Inc.

This Article grew out of a conference held at the UCLA School of Law in February 2010 under the auspices of PULSE. We thank the UCLA School of Law and the A. Barry & Lori Cappello Fund for funding this event. We also thank Forrest Havens and Julie Nicholson for helpful research assistance. Special thanks to Jerry Kang for his extraordinary mindmapping of the authors' working session during the conference. We would also like to thank the additional participants in the PULSE symposium, as well as the thoughtful commentators who have responded to this Article in this issue.

handwriting identification evidence in particular, do not currently possess—and absolutely must develop—a well-established scientific foundation. This can only be accomplished through the development of a research culture that permeates the entire field of forensic science. A research culture, we argue, must be grounded in the values of empiricism, transparency, and a commitment to an ongoing critical perspective. The forensic science disciplines need to substantially increase their commitment to evidence from empirical research as the basis for their conclusions. Sound research, rather than experience, training, and longstanding use, must become the central method by which assertions are justified. In this Article, we describe the underdeveloped research culture in the non-DNA forensic sciences, offer suggestions for how it might be improved, and explain why it matters.

INTRODUCTION

Long-used types of forensic science—fingerprint examination, handwriting analysis, firearms and toolmark comparison, and other forms of pattern and impression evidence—are mainstays of criminal prosecution. For roughly a hundred years, these comparison and identification methods have regularly and routinely been employed as legal evidence. For most of that period, courts, attorneys, jurors, and the public, as well as forensic analysts themselves, have largely accepted this evidence as trustworthy and uncontroversial.

In the last few years, the situation has changed dramatically. These methods and techniques now face more criticism and scrutiny than ever before. Latent fingerprint identification,[1] questioned document examination, and firearms comparison have been the targets of numerous admissibility challenges. Defendants have argued that this evidence is insufficiently valid to be admissible under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[2] and insufficiently accepted by the relevant scientific community to be admissible under *Frye v. United States*.[3] While most courts have continued to admit these forms of evidence, forensic practitioners have found themselves in the spotlight, forced to justify and defend practices that had previously evaded scrutiny.[4] Meanwhile, scandals involving crime laboratories have rippled across the nation: From Los Angeles to Charlotte, from Oklahoma City to Houston, stories of carelessness,

1. A latent fingerprint is an impression, invisible to the naked eye, left by a finger (or, more precisely, by friction ridge skin) on a surface. Latent prints are commonly recovered from crime scenes.

2. 509 U.S. 579 (1993). The Court, in an opinion by Justice Blackmun, held that the Federal Rules of Evidence required judges confronted with a challenge to scientific evidence to engage in a "flexible" inquiry whose "overarching subject is the scientific validity—and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission." *Id.* at 594–95. The Court elaborated on *Daubert*'s approach in *General Electric Co. v. Joiner*, 522 U.S. 136 (1997), which reaffirmed the Court's commitment to trial court gatekeeping and made clear that the appellate standard for review of the trial court's admissibility decisions was abuse of discretion. In *Kumho Tire Co. v. Carmichael*, the Court held that the district court's gatekeeping obligations extended to all forms of expert evidence and that judicial evaluation of reliability of expert evidence should focus on the particular task at issue in the specific case rather than the general validity of a field of expertise writ large. 526 U.S. 137, 141 (1999).

3. 293 F. 1013 (D.C. Cir. 1923). *Frye*'s key and oft-quoted language states:

> Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be *sufficiently established to have gained general acceptance in the particular field in which it belongs*.

Id. at 1014 (emphasis added). With any given forensic science, if the particular field, such as firearms comparison, is defined narrowly to consist only of firearms examiners, general acceptance cannot be in doubt. If the field is defined more broadly to include experts in all forms of pattern analysis, statisticians, and computer scientists, then the answer becomes less obvious. *See, e.g.*, 1 DANIEL L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 1:5, at 12–13 (2009–2010 ed.); DAVID H. KAYE, DAVID BERNSTEIN & JENNIFER L. MNOOKIN: THE NEW WIGMORE: EXPERT EVIDENCE § 6.3.3(b) (2d ed. 2010); Simon A. Cole, *Out of the* Daubert *Fire and Into the* Fryeing *Pan? Self-Validation, Meta-Expertise and the Admissibility of Latent Print Evidence in* Frye *Jurisdictions*, 9 MINN. J. L. SCI. & TECH. 453 (2008). In addition, many *Frye* states have inched towards a partial inquiry into validity. *See* 1 PAUL C. GIANNELLI & EDWARD L. IMWINKELRIED, JR., SCIENTIFIC EVIDENCE § 1.11, at 67 (2007); KAYE, BERNSTEIN & MNOOKIN, *supra*, § 7.4.2(b).

4. For a recent look at the variety of judicial reactions to these forms of evidence, see generally, KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 15; Jennifer L. Mnookin, *The Courts, the NAS, and the Future of Forensic Science*, 75 BROOK. L. REV. 1209 (2010). Handwriting evidence has received a more ambivalent reception than fingerprint identification or firearms comparison. *See* D. Michael Risinger, *Cases Involving the Reliability of Handwriting Identification Expertise Since the Decision in* Daubert, 43 TULSA L. REV. 477 (2007).

bias, incompetence, excessive coziness with prosecutors, and other embarrassing revelations have raised doubts about the trustworthiness and accuracy of some reported findings in a disturbing number of laboratories.[5] In 2004, the American fingerprint community faced its most high-profile fingerprint error ever when several highly experienced FBI examiners erroneously linked Oregon attorney (and Muslim convert) Brandon Mayfield to a fingerprint associated with the Madrid train bombing.[6] One study found that the trials of more than half of those defendants exonerated by post-conviction DNA testing included forensic evidence offered by the prosecution.[7] A follow-up study examining the underlying transcripts concluded that the testimony presented

5. Over the last twenty years, serious concerns have arisen in crime laboratories across the country, including in Boston, Chicago, Detroit, Houston, Los Angeles, New York, North Carolina, Oklahoma City, San Francisco, and West Virginia, as well as at the FBI laboratory. KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 1.4.1(a), at 22 & n.32. For examples from the voluminous press on these scandals, see JOHN F. KELLY & PHILLIP K. WEARNE, TAINTING EVIDENCE: INSIDE THE SCANDALS AT THE FBI CRIME LAB (1998); Tina Daunt, *LAPD Blames Faulty Training in DNA Snafu*, L.A. TIMES, July 31, 2002, at B3 (discussing the LAPD's accidental destruction of rape kits); Lianne Hart, *DNA Lab's Woes Cast Doubt on 68 Prison Terms*, L.A. TIMES, Mar. 31, 2003, at A19 (discussing the Houston crime lab scandal); Mandy Locke & Joseph Neff, *Inspectors Failed to Find SBI Faults*, CHARLOTTE OBSERVER, Aug. 26, 2010, *available at* http://www.charlotteobserver.com/2010/08/26/1643668/inspectors-failed-to-find-sbi.html; Moises Mendoza, *HPD Fingerprinting Trouble Not Unique*, HOUSTON CHRON., Dec. 13 2009, at B1 (giving context to Houston fingerprint lab problems); Steve Mills & Maurice Possley, *Report Alleges Crime Lab Fraud: Scientist Is Accused of Providing False Testimony*, CHI. TRIB., Jan. 14, 2001, at C1 (discussing the Pamela Fish scandal in Chicago); Maggie Mulvihill & Franci Richardson, *Unfit Cops Put in Key Evidence Unit; Fingerprint Handlers Were All Thumbs*, BOSTON HERALD, May 7, 2004, at 2; Maurice Possley, Steve Mills & Flynn McRoberts, *Scandal Touches Even Elite Labs: Flawed Work, Resistance to Scrutiny Seen Across U.S.*, CHI. TRIB., Oct. 21, 2004, at C1; Jonathan Saltzman & John R. Ellement, *Mass. DNA Lab's Lapses Draw Beacon Hill Inquiry: Delays, Errors Laid to Lack of Oversight*, BOSTON GLOBE, Jan. 17, 2007, at 1A (discussing the Massachusetts state crime lab scandal); Ben Schmitt & Joe Swickard, *Troubled Detroit Police Crime Lab Shuttered: State Police Audit Results 'Appalling,' Wayne County Prosecutor Declares*, DETROIT FREE PRESS, Sept. 26, 2008, at 1 (discussing the multiple problems that led to the Detroit crime lab's closure); Jaxon Van Derbeken, *SFPD Drug-Test Technician Accused of Skimming*, S.F. CHRON., Mar. 10, 2010, at A1 (discussing the San Francisco crime lab scandal); Murray Weiss, *Criminal Errors*, N.Y. POST, Dec. 4, 2007 (discussing a scandal at an NYPD crime lab); Jim Yardley, *Inquiry Focuses on Scientist Employed by Prosecutors*, N.Y. TIMES, May 2, 2001, at 2 (discussing the Joyce Gilchrist Oklahoma scandal); *Court Invalidates a Decade of Blood Test Results in Criminal Cases*, N.Y. TIMES, Nov. 12, 1993, at A20 (discussing the Fred Zain West Virginia scandal). The problems that have come to light have occurred in a variety of forensic areas, including serology, bloodstain pattern analysis, DNA, fingerprint identification, and others. While our primary focus in this Article is on pattern evidence, these scandals serve as a reminder that the issues we describe warrant thoughtful attention throughout forensic science, not just in the pattern identification arena.

6. *See* Robert B. Stacey, *Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case*, 54 J. FORENSIC IDENTIFICATION 706 (2004); OVERSIGHT & REVIEW DIV., U.S. DEP'T OF JUSTICE, A REVIEW OF THE FBI'S HANDLING OF THE BRANDON MAYFIELD CASE (2006), *available at* http://www.usdoj.gov/oig/special/s0601/PDF_list.htm.

7. Brandon L. Garrett, *Judging Innocence*, 108 COLUM. L. REV. 55, 82 (2008).

by forensic analysts had frequently been overstated or misleading.[8] While currently available information does not permit quantification beyond the sample of cases examined, these studies do suggest that misleading or erroneous forensic science has contributed to a substantial number of false convictions. A number of academics began to examine the research foundation of some long-used forensic disciplines and found that claims were frequently supported by far less rigorous research than might have been expected. And in February 2009, the prestigious National Academy of Sciences (NAS) issued a long-awaited report on the forensic sciences that concluded: "With the exception of DNA analysis, . . . no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source."[9]

The NAS Report suggested a number of major improvements for forensic science. Most significantly, it called for the creation of an entirely new, independent federal agency to oversee and regulate the practices of forensic science, and to ensure the development of rigorous research to determine the capabilities and the limits of forensic science.[10] This combination of events—legal challenges, laboratory scandals, widely publicized errors, skeptical scholarship, and a highly critical national report—has focused sometimes unwelcome but badly needed attention on the forensic sciences. These developments offer the opportunity for reflection and improvement.

---

8. *See* Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 VA. L. REV. 1 (2009). There are three important caveats to be made regarding this article's conclusions. First, the single most problematic form of evidence in the Garrett & Neufeld study was microscopic hair analysis, which is now typically used as an adjunct to mitochondrial DNA assessment of hair. Some have therefore argued that this makes the conclusions from Garrett's original study (and its follow-up) largely moot. *See, e.g.*, JOHN COLLINS & JAY JARVIS, CRIME LAB REPORT, THE WRONGFUL CONVICTION OF FORENSIC SCIENCE (2008), *available at* http://www.crimelabreport.com/library/pdf/wrongful_conviction.pdf. However, there is no reason to believe that the culture that produced these frequent overstatements and failures to adhere strictly to conclusions warranted by the evidence was limited to microscopic hair analysis. Second, it is important to recognize that some of the expert testimony was not erroneous or overstated, even if it turned out to invite an incorrect inference about the identity of the perpetrator. *See, e.g.*, COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCIS. CMTY., NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 120 (2009) [hereinafter NAS REPORT] (pointing out that evidence such as physical inspection of hair or paint that merely identifies a trace as falling into a large class of potential sources is accurate even if it turns out that the defendant is not the source). Third, of course, we have virtually no direct information in these cases about how the jury perceived the forensic science evidence. It would therefore be dangerous to infer from the mere fact of conviction that the jury found the forensic science evidence either persuasive or critical in any given case; however, it would be equally questionable to presume that it did not. The prosecution proffered it, after all, to aid in conviction.

9. NAS REPORT, *supra* note 8, at 7.

10. *Id.*

Now, roughly two years after the publication of this major report, where do the pattern and impression disciplines and the forensic sciences more generally stand?[11] What are the ongoing problems in these fields? What ought to be the intellectually significant and yet practically realizable goals for improving forensic science evidence over the next decade or two? The purpose of this Article is to describe what we think forensic science most needs in order to best serve justice, the legal system, the public, and its own practitioners. Our central argument is that the pattern identification disciplines, and forensic science more generally, do not currently possess—and absolutely must develop—an adequate research culture. In what follows, we will outline the essential elements of a research culture, provide examples to support our claim that within these disciplines such a culture is weak or faltering, and offer some concrete suggestions for how a research culture might be created.

The authors of this Article are a diverse group. This group includes those who are quite regularly labeled critics of forensic science, as well as defenders, including some who toil in the fields of forensic science every day. It includes forensic analysts from several fields who regularly appear in court testifying to the reliability of forensic evidence, as well as those who have appeared in court criticizing such evidence. Some of us are pursuing empirical research about forensic science; others write more conceptually about its strengths and limitations; still others among us spend more time practicing forensic science than writing about it. One of us is a former director of a major crime laboratory. The academics in this group come from multiple fields and varying disciplinary backgrounds, including law, cognitive psychology, chemistry, molecular biology, forensic science, and the sociology of science. One member of the group was on the NAS Committee and helped write its report.

Given the breadth of backgrounds, disciplines, and points of view, and given the current controversies surrounding forensic science, it will come as

11. Our primary focus is on pattern and impression evidence. These disciplines include fingerprint analysis, firearms and toolmark comparison, questioned document examination, shoeprint examination, microscopic hair comparison, tire tread comparison, blood spatter analysis, bite mark analysis, and other physical object comparisons. These disciplines have in common that they attempt to determine whether or not a particular pattern or impression—be it a shoeprint, a tire tread, a fingerprint, or a bullet—can be associated with a particular source. (Blood spatter analysis is an exception, as it attempts to use the pattern of blood to infer something about the physical events that gave rise to it). Although we focus primarily on pattern evidence, many of our arguments apply to forensic science more broadly. Tracing out with specificity where they do and do not fully apply across the broader range of forensic sciences—from DNA analysis to arson investigation to toxicology—is beyond the scope of this Article. We recognize that different portions of the forensic science landscape vary in the extent to which they already possess a robust research culture, but we believe that the forensic science enterprise, as a whole, would benefit from more focused efforts to develop the outlook and practices referred to in this Article as a research culture.

no surprise that this diverse group of authors does not agree about everything. We cannot pretend to share a wholly unified vision for the future of forensic science. But what is striking—and what generated this Article—is that there is a good deal about which we do agree. The purpose of this Article is to focus on these substantial areas of agreement. We aim to lay out our shared understanding of some of the current problems in forensic science, and our consensus on how to improve the pattern identification fields, and the rest of the forensic science enterprise, so that they will rest on an appropriately secure foundation as they continue to provide valuable evidence to the criminal justice system into the twenty-first century.

Significantly, despite our diverse backgrounds and points of view, we agree on many aspects of both the diagnosis of current difficulties and a direction for a cure. In our collective opinion, the pattern identification disciplines, as well as other forms of forensic science evidence, must be placed on a more rigorous scientific foundation. More generally, we believe that a significant culture shift is required: Forensic science needs to focus more on science than on law, to shift from a quasi-adversarial perspective to a research orientation. In short, we call for the development and instantiation of what we will term a *research culture* within forensic science. The emergence of a research culture would affect how evidence is understood, change analysts' relationship to empirical data, and alter how evidence is reported. We do not delude ourselves that change comes easily or that a culture shift alone will immediately ensure that all forensic analyses are well founded.[12] But we believe that this transition is both necessary and, while difficult, genuinely feasible.

In what follows, we begin with a brief overview of the NAS Report, a watershed publication for the assessment of the current state of the forensic sciences. We use this report as a springboard to describe our consensus about what the forensic sciences need most in order to attain a solid footing over the next decades: to wit, the creation of a robust research culture, in which empirical evidence and careful scrutiny regarding the evidentiary warrant

---

12. To be sure, scientists steeped in the research culture we describe in this Article also sometimes make claims that outstrip their data or promote methods before the application has been shown to be fully robust or before all its limitations are clear. *See, e.g.*, DAVID H. KAYE, THE DOUBLE HELIX AND THE LAW OF EVIDENCE 51–53 (2010) (describing the early claims of the developer of DNA profiling). Nevertheless, in a competitive research culture, any premature enthusiasm or dubious assertions are likely to be met with criticism from others in the community, leading in the long run to a much more secure foundation for the applications of the theory or procedure. *See id.* at 53–54, 117, 119–20, 123–26, 138 (describing how the scientists who promoted or defended DNA identification responded to various published criticisms); *see also* Jennifer L. Mnookin, People v. Castro: *Challenging the Forensic Use of DNA Evidence*, *in* EVIDENCE STORIES 207 (Richard Lempert ed., 2006). *See generally* D. Michael Risinger, *The Irrelevance, and Central Relevance, of the Boundary Between Science and Non-Science in the Evaluation of Expert Witness Reliability*, 52 VILL. L. REV. 679, 700–12 (2007).

for whatever claims are made become part of the ordinary way of thinking about forensic practices. In Part II, we describe what we see as the critical components of a research culture, including a focus on empirical evidence, transparency, and a consistently critical and reflective perspective on claims of knowledge. In Part III, we provide a number of examples and illustrations to show why we do not believe that the research culture within forensic sciences, and within pattern and impression evidence in particular, is presently either well developed or robust. In the final Part, we offer a variety of suggestions, some of them familiar and some of them more innovative, for creating and fostering a research culture for forensic science.

I. THE NATIONAL ACADEMY OF SCIENCES REPORT

In February 2009, The National Academy of Sciences issued its major report on forensic science.[13] Congress commissioned the report late in 2005 at the behest of the forensic science community itself. The Academy appointed a panel of judges, scholars, and forensic and legal practitioners to write the report. This committee heard more than sixteen days of testimony—more than eighty witnesses in eight meetings over a two-year period—from a variety of leading forensic scientists and academic researchers.[14]

In addition to the major recommendation to create a National Institute of Forensic Science (NIFS), the committee put forward two other important structural recommendations: the removal of public forensic science laboratories from administrative control of law enforcement agencies or prosecutors' offices;[15] and the gradual abolition of state and local coroners' offices in favor of a medical examiner system.[16] Given the committee's key finding that an inadequate research basis existed for claims often made in forensic science, most of the recommendations were concerned with improving the science in forensic science. Specific recommendations included, among others:

> (1) Foundational research that would assess the validity and reliability of methods used in the analysis of evidence, especially pattern evidence.[17]

13. NAS REPORT, *supra* note 8.
14. Harry T. Edwards, Co-Chair, Committee on Identifying the Needs of the Forensic Science Community, The National Academy of Sciences Report on Forensic Sciences: What It Means for the Bench and Bar (May 6, 2010). For a close look at the committee and who was on it, see D. Michael Risinger, *The NAS/NRC Report on Forensic Science: A Path Forward Fraught With Pitfalls*, 2010 UTAH L. REV. 225.
15. NAS REPORT, *supra* note 8, at 23–24.
16. *Id.* at 29.
17. *Id.* at 22–23.

(2) Further research into the issues of cognitive bias and its effects on forensic decisionmaking. The committee recognized the significant need to investigate when contextual or confirmational bias might affect examiners' processes or their conclusions, and the need both to study its extent and to develop countermeasures.[18]

(3) Standardization of laboratory reports and a standard definition of terms, especially those expressing the association between an item of evidence and a possible source.[19]

(4) Mandatory accreditation of all forensic science laboratories that process evidence for court and mandatory certification of all forensic scientists who analyze evidence.[20]

(5) A mandatory code of ethics that is tied to certification and makes possible the removal of serious ethical violators from the practice of forensic science.[21]

The NAS Committee was not charged specifically with examining the issues surrounding pattern and impression evidence, although the final report does emphasize these areas. This focus emerged as the committee heard testimony about the present state of research and the validity and reliability of forensic science methods. In testimony presented, various types of pattern evidence were cited as poster children for the lack of scientific foundation in forensic science and the need for more research to establish the validity (or lack of it) in forensic science's analytical methods.[22] Fairly or not, the report reflects this emphasis, and in this Article, we too focus primarily on the pattern and impression areas. Nevertheless, it is important to recognize that the pattern evidence areas are not alone in generating the concerns expressed in the NAS Report or in this Article. Every area of forensic science, including DNA typing, described by the NAS Committee as the "gold standard" of forensic science, suffers to some degree from the problems the report ascribed to pattern evidence. Most of the recommendations in the NAS Report are global in their reach; they are intended to apply to forensic science as a whole. Similarly, although we focus on pattern and impression evidence processed by human analysts using visual examination, many of our arguments apply beyond these domains. At the

18. *Id.* at 24.
19. *Id.* at 21.
20. *Id.* at 25.
21. *Id.* at 26.
22. Chapter five of the NAS Report discusses and offers summary assessments of, for example, biological evidence, analysis of controlled substances, friction ridge analysis, shoeprints and tire tracks, toolmark and firearms identification, hair and fiber evidence, questioned document examination, paint and coating evidence, arson and explosives evidence, bitemark analysis, and bloodstain pattern analysis. *See id.* at 127–83. Chapter nine focuses entirely on medicolegal death analysis. *See id.* at 241–68.

same time, we recognize that forensic science culture is not monolithic or unitary. We hope that our remarks in the context of pattern and impression evidence will encourage further discussion and attention to the question of how to create, develop, or improve the research culture in other areas, including forensic chemistry, DNA analysis, fire investigation, and medicolegal death investigation.

We all agree that publication of the report was a watershed moment for the forensic sciences. The report continues to generate both attention and controversy. Already it has prompted, or at least spurred, some degree of change in forensic science practice.[23] It continues to influence practicing forensic scientists themselves, as well as those who interact with forensic disciplines, including lawyers and judges, government officials, and government regulatory and funding entities, such as the National Institute of Justice (NIJ), National Institute of Standards and Technology (NIST), and the National Science Foundation (NSF).[24]

We also agree with much of the content of the report.[25] Perhaps most significantly, we agree with the National Academy's central and important conclusion that the traditional forensic sciences are, at this point, inadequately supported by empirical data that would justify the strong claims analysts frequently make. We believe that numerous assertions made both in routine practice and in court are neither backed up by sufficient empirical data or

23. *See, e.g.*, INT'L ASS'N FOR IDENTIFICATION, IAI RESOLUTION 2010–18, at 1 (2010) [hereinafter IAI RESOLUTION], *available at* http://onin.com/fp/IAI_resolution_2010-18.pdf (reflecting a "change [in] the official position of the Association related to Friction Ridge Examinations based on advances in the science and scientific research").

24. The National Science Foundation recently funded a workshop at Northwestern Law School called "Cognitive Bias and Forensic Science" largely designed to encourage social and cognitive psychologists to conduct empirical studies to improve our understanding of factors that affect forensic science judgments and decisions. Similarly, the National Institute of Justice has solicited research proposals from social scientists to study, for example, "'context bias' and the need for a greater understanding of the scope of this issue in forensic laboratories." NAT'L INST. OF JUSTICE, U.S. DEP'T OF JUSTICE, SOLICITATION: SOCIAL SCIENCE RESEARCH IN FORENSIC SCIENCE 3 (2010).

25. One author of this Article served on the NAS Committee, and several others gave invited presentations to the committee. Nonetheless, we do not agree with every sentence or every detail of every argument in this report. Certainly we each have both nits to pick as well as admiration for its strengths; indeed, several of us have already expressed both our criticism and our praise in print. *See, e.g.*, Simon A. Cole, *Who Speaks for Science? A Response to the National Academy of Sciences Report on Forensic Science*, 9 LAW, PROBABILITY & RISK 25 (2010); Itiel E. Dror, *How Can Francis Bacon Help Forensic Science? The Four Idols of Human Biases*, 50 JURIMETRICS J. 93 (2009); David H. Kaye, *The Good, the Bad, and the Ugly: The NRC Report on Strengthening Forensic Science in America*, 50 SCI. & JUST. 8 (2010); Jonathan J. Koehler, *Forensic Science Reform in the 21st Century: A Major Conference, a Blockbuster Report and Reasons to Be Pessimistic*, 9 LAW, PROBABILITY & RISK 1 (2010); Mnookin, *supra* note 4; D. Michael Risinger, *The NAS/NRC Report on Forensic Science: A Glass Nine-Tenths Full (This Is About the Other Tenth)*, 50 JURIMETRICS J. 21 (2009); Risinger, *supra* note 14. For a collection of responses from the forensic science community, see generally *id*.

research, nor are the kinds of claims that can be justified or validated simply by reference to longstanding experience.[26] We have in mind, for example, asserting an error rate of zero for the methodology of latent fingerprint identification; testifying that forensic practitioners have an adequate empirical and experiential basis for confidently determining in run-of-the-mill cases that two prints—or shoe marks or firearms or handwriting exemplars—share a common source to the exclusion of all other possible sources;[27] claiming confidence based on experience that analysts have taken adequate steps to counter the effects of bias and context; or averring that the techniques used by forensic pattern disciplines follow "the scientific method."[28]

We do recognize that experience, training, and longstanding investigatory and legal use can be sources of legitimate knowledge for pattern identification analysts. We also recognize that experience and training can give examiners,

---

26. *See, e.g.*, Mark A. Acree, *What Is Science? The Dilemma of Fingerprint Science Revisited*, 14 PRINT 4 (1998); Michelle Reznicek et al., *ACE-V and the Scientific Method*, 60 J. FORENSIC IDENTIFICATION 87 (2010); Michele Triplett & Lauren Cooney, *The Etiology of ACE-V and Its Proper Use: An Exploration of the Relationship Between ACE-V and the Scientific Method of Hypothesis Testing*, 56 J. FORENSIC IDENTIFICATION 345 (2006); Norah Rudin & Keith Inman, *The Experience Fallacy*, CACNEWS, Fourth Quarter 2010, at 10, *available at* http://cacnews.org/news/4thq10.pdf.

27. There may be rare instances when the strong claim of individualization could be warranted because both sides agree that the universe of potential suspects is small—such as, for example, in a situation where it is uncontested that a murder was committed by one of a small group of people in a locked house. But this inference is warranted in these circumstances because of the reduced size of the possible suspect population, not because of the prints' power to individualize to the exclusion of all others in the universe. *See* KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 15.2; David H. Kaye, *Probability, Individualization, and Uniqueness in Forensic Evidence: Listening to the Academics*, 75 BROOK. L. REV. 1163, 1174–75 (2010) (arguing that testing most, but not all, of a closed set of suspects can justify the conclusion that the trace evidence originated from a single individual).

28. To some extent, these specific rhetorical claims are being modified in the aftermath of, and in response to, the report. Indeed, the day after the report was issued, the president of the International Association of Identification (IAI) wrote the membership: "It is suggested that members not assert 100% infallibility (zero error rate) when addressing the reliability of fingerprint comparisons." Memorandum From Robert Garrett, President, Int'l Ass'n for Identification, to the Membership of Int'l Ass'n for Identification (Feb. 19, 2009), *available at* http://www.theiai.org/current_affairs/nas_memo_20090219.pdf. In July 2010, the IAI also opened the door to probabilistic testimony regarding the likelihood of a fingerprint match and rescinded a 1979 resolution that limited such testimony to only three possible conclusions: individualization, exclusion, and unknown. For the recent resolution, see IAI RESOLUTION, *supra* note 23. In the summer of 2010, the chairman of the Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), a standard-setting organization for friction ridge analysis, issued a "clarification" asserting that the phrase "to the exclusion of all others" is likely to be removed from its Friction Ridge Examination Methodology materials. Letter From Leonard G. Butt, Chairman, Scientific Working Group on Friction Ridge Analysis, Study, and Tech., to Whom It May Concern (June 29, 2010), *available at* http://www.swgfast.org/Comments-Positions/SWGFAST_NAS_Position_Clarification.doc. While we believe that all of these terminological shifts are positive developments, they do not negate or eliminate our more general arguments about the continued lack of a research culture in much of the pattern identification sciences, nor do they solve the problem of how to responsibly characterize the probative value of the results of an analysis. On this latter difficulty, see generally KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 15; Mnookin, *supra* note 4.

from a subjective point of view, sincere and deeply held confidence about their ability to do what they claim to do. But we do not believe that experience and training alone can validate universalist claims, such as the claim that latent fingerprint identification experts can individualize the source of a print to the exclusion of all other possibilities,[29] or the claim that document examiners' experience enables them to assess the entire range of differences between two handwriting exemplars that can still be consistent with authorship by the same hand.[30]

More generally, we believe that not enough is yet known about a significant range of important questions. Consider, for example, the following: Precisely what are the capabilities and limitations of any particular pattern discipline? How often do pattern identification analysts make mistakes? When these errors occur, what causes them? How should error be defined,[31] and what circumstances tend to increase the risk of error? How prevalent is the effect of cognitive bias on the activity of forensic examiners? When might access to contextual information affect forensic examiners' cognitive processes, or even their final determinations? How frequently might a portion of two fingerprints—or striation marks on bullets, or toolmarks, or handwriting specimens—share any given degree of similarity even if they derive from different sources? How does the use of large databases or new imaging technologies help these disciplines, and what dangers may new technologies pose? Just how much visual information is sufficient to undertake an accurate analysis of a handwriting specimen, a latent fingerprint, or a firearm? To what extent does training improve examiner accuracy? What kind of training is most effective?

---

29. For discussion of the dubious underpinnings of assertions of "global individualization," see KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 15.

30. *See, e.g.*, *What Is Forensic Document Examination?: Handwriting Examination*, SE. ASS'N OF FORENSIC DOCUMENT EXAMINERS, http://www.safde.org/whatwedo.htm (last visited Jan. 28, 2011):

> Handwriting identification is based on the principle that, while handwriting within a language tends to be alike to the degree that we can meaningfully read it, there are individual features that distinguish one person's writing from that of another. Just as no two people are exactly alike, the handwritings of no two people are exactly alike in their combination of characteristics. There are, of course, natural variations *within* the handwriting of each individual. These variations must be closely and carefully studied by the examiner, so that he can distinguish between what is a "variation" and what is a "difference."
>
> The examiner must also be cognizant of the differences between "class characteristics" and "individual characteristics." Class characteristics are those which are common to a group such as a particular writing system, family grouping, foreign language system, or professional group. Individual characteristics are those which are personal or peculiar letters or letter combinations, which, taken together, would not occur in the writing of another person.

31. *See generally* KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 7.3.2(c); D. Michael Risinger, *Whose Fault?:* Daubert*, the NAS Report and the Notion of Error in Forensic Science*, 49 FORDHAM URB. L.J. 519 (2010).

The honest response to all of these questions is that we do not yet know. Suggestive research is emerging in some of these areas, including contributions from several co-authors of this Article.[32] But we all agree that as of yet, the research basis that would permit a satisfying scientific answer to any of the above questions does not exist.

To be sure, we also recognize that the absence of evidence is not necessarily evidence of absence.[33] Until recently, virtually no institution—not the courts,

---

32. *See*, *e.g.*, Jan Beck, *Sources of Error in Forensic Handwriting Evaluation*, 40 J. FORENSIC SCI. 78 (1995); Silvia Bozza et al., *Probabilistic Evaluation of Handwriting Evidence: Likelihood Ratio for Authorship*, 57 J. ROYAL STAT. SOC'Y SERIES C: APPLIED STAT. 329 (2008); Stephen G. Bunch, *Consecutive Matching Striation Criteria: A General Critique*, 45 J. FORENSIC SCI. 955 (2000); Christopher Champed, *Edmond Locard—Numerical Standards & "Probable" Identifications*, 45 J. FORENSIC IDENTIFICATION 136 (1995); Christophe Champod et al., *Establishing the Most Appropriate Databases for Addressing Source Level Propositions*, 44 SCI. & JUST. 153 (2004); Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 FORENSIC SCI. INT'L 74 (2006); Itiel E. Dror & Robert Rosenthal, *Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts*, 53 J. FORENSIC SCI. 900 (2008); Itiel E. Dror & David Charlton, *Why Experts Make Errors*, 56 J. FORENSIC IDENTIFICATION 600 (2006); Adrian G. Dyer et al., *An Insight Into Forensic Document Examiner Expertise for Discriminating Between Forged and Disguised Signatures*, 53 J. FORENSIC SCI. 1154 (2008); Bryan Found & Doug Rogers, *The Probative Character of Forensic Handwriting Examiners' Identification and Elimination Opinions on Questioned Signatures*, 178 FORENSIC SCI. INT'L 54 (2008); Moshe Kam et al., *Signature Authentication by Forensic Document Examiners*, 46 J. FORENSIC SCI. 884 (2001); Moshe Kam & Erwei Lin, *Writer Identification Using Hand-Printed and Non-Hand-Printed Questioned Documents*, 48 J. FORENSIC SCI. 1391 (2003); Glenn Langenburg, *A Performance Study of the ACE-V Process: A Pilot Study to Measure the Accuracy, Precision, Reproducibility, Repeatability, and Biasability of Conclusion Resulting From the ACE-V Process*, 59 J. FORENSIC IDENTIFICATION 219 (2009); Glenn Langenburg et al., *Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology When Conducting Fingerprint Comparisons*, 54 J. FORENSIC SCI. 571 (2009); D. Meuwly, *Forensic Individualisation From Biometric Data*, 46 SCI. & JUST. 205 (2006); Cedric Neumann et al., *Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Any Number of Minutiae*, 52 J. FORENSIC SCI. 54 (2007); Beatrice Schiffer & Christophe Champod, *The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingermark Individualisation*, 167 FORENSIC SCI. INT'L 116 (2007); Sargur N. Srihari et al., *Discriminability of Fingerprints of Twins*, 58 J. FORENSIC IDENTIFICATION 109 (2008); Sargur N. Srihari et al., *Individuality of Handwriting*, 47 J. FORENSIC SCI. 856 (2002); Kasey Wertheim et al., *A Report of Latent Print Examiner Accuracy During Comparison Training Exercises*, 56 J. FORENSIC IDENTIFICATION 55 (2006); Steve Gutowski, *Error Rates in Fingerprint Examination: The View in 2006*, FORENSIC BULL. (Nat'l Inst. of Forensic Sci., Austl.), Autumn 2006, at 18. This list is not exhaustive and should not be considered an endorsement of particular studies. While some of the research listed above is first rate, other studies may have significant flaws. But these works do at least constitute efforts to examine empirically relevant questions that have often been assumed rather than investigated.

33. Carl Sagan appears to have originated the felicitous phrasing "absence of evidence is not evidence of absence." CARL SAGAN, THE DEMON-HAUNTED WORLD: SCIENCE AS A CANDLE IN THE DARK 213 (1996). The difficulty with this aphorism is that the absence of evidence supporting a theory following a search for it can be evidence of the falsity of the theory. *Cf.* Elliott Sober, *Absence of Evidence and Evidence of Absence: Evidential Transitivity in Connection With Fossils, Fishing, Fine-tuning, and Firing Squads*, 143 PHIL. STUD. 63 (2009). However, read in context, Sagan was criticizing as "impatience with ambiguity" both the notion that whatever has not been proved false must be true and the opposite, that what has not been proved true must be false. SAGAN, *supra*, at 213. Inferring validity from the

not government funding agencies, very few research organizations or forensic science laboratories—was investing a great deal of time, energy, or resources into answering these questions.[34] We therefore lack any major body of published scientific research directed at empirically validating the conceptual claims and underpinnings of the pattern identification forensic disciplines.[35] As a body of research continues to emerge, and we learn more about the frequency and types of errors that do occur, we may well find that many current practices turn out to have tolerably low error rates. As we develop and validate methods for probabilistic assessments of fingerprints, documents, or firearms, we may learn that in many circumstances the chances of a coincidental match are extremely remote, and we will certainly learn more about how common or remote they truly are. It could turn out that analysts' experience-based intuitive judgments about the correspondence sufficient to declare a match, even if not presently quantified or formally specified, are generally quite accurate. It may be that the biasing effects of access to contextual information extraneous to the forensic analysis rarely impact an examiner's conclusion or ultimate judgment when the information contained within the pattern is sufficiently clear.

All of this is possible. But none of it is adequately established yet. While our collective hunches about what the expanding pool of research will reveal

---

fact that many kinds of forensic science have not been proved invalid, and inferring invalidity from the lack of scientific proof of validity are both dangers to avoid.

34. There have been, to be sure, individuals engaged in some degree of research. *See supra* note 32. At the institutional level, there are also limited exceptions to these generalizations: The present research efforts emerging from the University of Lausanne and the period at the University of California, Berkeley in which several students under the tutelage of chemist and forensic scientist Paul Kirk pursued fundamental research in forensics are perhaps the most notable. On the current research program at Lausanne, see, for example, *School of Criminal Justice (ESC)*, UNIVERSITÉ DE LAUSANNE, http://www.unil.ch/central/page2904_en.html (last visited Jan. 29, 2011).

35. NAS REPORT, *supra* note 8, at 107–08 ("Much forensic evidence—including, for example, bite marks and firearm and toolmark 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."). There are, to be sure, a handful of researchers beginning to change this, but a substantial body of work points out the absence of an adequate scientific foundation or empirical basis for the forensic sciences. *See, e.g.*, SIMON A. COLE, SUSPECT IDENTITIES: A HISTORY OF FINGERPRINTING AND CRIMINAL IDENTIFICATION (2001); Simon A. Cole, *Is Fingerprint Identification Valid? Rhetorics of Reliability in Fingerprint Proponents' Discourse*, 28 LAW & POL'Y 109 (2006); David L. Faigman, *Anecdotal Forensics, Phrenology, and Other Abject Lessons From the History of Science*, 59 HASTINGS L.J. 979 (2008); Jennifer L. Mnookin, *Fingerprint Evidence in an Age of DNA Profiling*, 67 BROOK. L. REV. 13 (2001) [hereinafter Mnookin, *Fingerprint Evidence*]; Jennifer L. Mnookin, *The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate*, 7 LAW, PROBABILITY & RISK 127 (2008) [hereinafter Mnookin, *Confessions*]; D. Michael Risinger et al., *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise"*, 137 U. PA. L. REV. 731 (1989); Michael J. Saks, *The Legal and Scientific Evaluation of Forensic Science (Especially Fingerprint Expert Testimony)*, 33 SETON HALL L. REV. 1167, 1186–87 (2003); David A. Stoney, *Measurement of Fingerprint Individuality*, *in* ADVANCES IN FINGERPRINT TECHNOLOGY 327, 329 (Henry C. Lee & R.E. Gaensslen eds., 2d ed. 2001).

vary, we all expect that additional, high-quality research will confirm that many forensic science techniques, including many kinds of pattern and impression evidence, do have a considerable degree of discriminatory power and that there exists significant variation in discriminatory power across fields and within any given field, depending upon particularized circumstances.[36] Furthermore, we all agree that we presently lack sufficient knowledge regarding the precise extent of this power or its limits.

Calling, therefore, for more research into these important questions, is both obvious and necessary. Here, again, we largely agree with the NAS Report's conclusion:

> In most areas of forensic science, no well-defined system exists for determining error rates, and proficiency testing shows that some examiners perform poorly . . . .
>
> In most forensic science disciplines, no studies have been conducted of large populations to establish the uniqueness of marks or features. Yet, despite the lack of a statistical foundation, examiners make probabilistic claims based on their experience. A statistical framework that allows quantification of these claims is greatly needed. These disciplines also critically need to standardize and clarify the terminology used in reporting and testifying about the results and in providing more information.
>
> Little rigorous research has been done to validate the basic premises and techniques in a number of forensic science disciplines. The committee sees no evident reason why conducting such research is not feasible . . . .[37]

To be sure, calling for more research is hardly a radical or controversial suggestion. Indeed, in the aftermath of the NAS Report, calls for more research have been widespread. Despite the report's contentious reception, and notwithstanding the significant disagreements within forensic science, we cannot actually point to anyone who has argued that more research, in the abstract, is a bad idea. We have certainly heard it said that more research is not needed

36. We all would predict, for example, that latent fingerprint identification will turn out to have a good deal more discriminatory power across a broader range of circumstances than forensic odontology (bitemark analysis). On bitemarks, see Mary A. Bush et al., *Statistical Evidence for the Similarity of the Human Dentition*, 56 J. FORENSIC SCI. (forthcoming 2011); D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99 (2000); Michael J. Saks, *Merlin & Solomon: Lessons From the Law's Formative Encounters With Forensic Identification Science*, 49 HASTINGS L.J. 1069 (1998). On the need to focus on the particularized task at hand rather than making global, field-wide admissibility judgments, see generally, D. Michael Risinger, *Defining the "Task at Hand": Non-Science Forensic Science After* Kumho Tire v. Carmichael, 57 WASH. & LEE L. REV. 767 (2000).

37. NAS REPORT, *supra* 8, at 188–89.

for basic validation.[38] We have heard it said that the costs of research need to be balanced against other needs. And we have heard it said that forensic science laboratories are too busy to undertake, or even participate, in research. While not everyone views more research as imperative, we are not aware of anyone who, in print, or even in the hallways of conferences or crime labs, opposes the very idea of research in the abstract.

But "more research," imprecisely defined, is not enough. What forensic science needs is the creation and institutionalization of a research culture.

II. WHAT IS A RESEARCH CULTURE?

What do we mean by a research culture? Put simply, we mean a culture in which the question of the relationship between research-based knowledge and laboratory practices is both foregrounded and central. We mean a culture in which the following questions are primary: What do we know? How do we know that? How sure are we about that? We mean a culture in which these questions are answered by reference to data, to published studies, and to publicly accessible materials, rather than primarily by reference to experience or craft knowledge, or simply assumed to be true because they have long been assumed to be true.

Before elaborating on the meaning of a research culture—and before presenting examples of the absence of a deep and robust research culture within forensic science together with suggestions for how to build it—it is critical to make one point: While we firmly believe that a research culture needs to become both more central and more entrenched within forensic science more generally, and within the pattern and impression disciplines specifically, this does not—and should not—mean that all forensic practitioners should henceforth

38. *See* AFTE Comm. for the Advancement of the Sci. of Firearm and Tool Mark Identification, *The Response of the Association of Firearm and Tool Mark Examiners to the February 2009 National Academy of Science Report "Strengthening Forensic Science in the United States: A Path Forward"*, 41 AFTE J. 204, 205 (2009) ("There is an extensive body of research, extending back over one hundred years, which establishes the accuracy, reliability, and validity of conclusions rendered in the field of firearm and toolmark identification."); Jeffrey G. Barnes, *History*, *in* THE FINGERPRINT SOURCEBOOK 1–17 (2010), *available at* http://www.ncjrs.gov/pdffiles1/nij/225321.pdf ("Study, research, and experimentation have led to and supported fingerprints as a means of individualization and a forensic tool of incalculable value. The research and practical knowledge accumulated over the course of many centuries well support the science."); Memorandum From Robert J. Garrett, *supra* note 28 ("There is no research to suggest that properly trained and professionally guided examiners cannot reliably identify whole or partial fingerprint impressions to the person from whom they originated."); *SWGGUN Systemic Requirements/Recommendations for the Forensic Firearm and Toolmark Laboratory*, SCIENTIFIC WORKING GROUP FOR FIREARMS & TOOLMARKS, http://www.swggun.org/guidelinedocs/SWGGUN%20Systemic%20Requirements.pdf (last modified Apr. 23, 2010) ("The reliability of the science has been demonstrated and supported through proficiency tests and validity studies over many decades.").

be doing research. To the contrary. Even with a research culture in place, most forensic practitioners will continue simply to practice forensic science. Some forensic practitioners might be "test subjects" for researchers—the objects of research study to help achieve a better understanding of the strengths and limitations of their methods and techniques. They might sometimes partner with researchers to develop projects, or to evaluate the practical feasibility of a given research design; on other occasions, they might assess what research needs they deem especially significant. But even these practitioners need not, and indeed often should not, be the primary producers of the research themselves.[39]

Medicine provides an instructive analogy. Modern medicine is a product of both craft knowledge and structured research. Whether medicine incorporates more of a research culture than forensic science is perhaps debatable, but certainly evidence-based medicine coexists with a more experience-based, clinical practice orientation still widely influential among doctors.[40] The point for our purposes, however, is that many more physicians make use of research than produce it. Some physicians certainly do pursue research alongside clinical practice,[41] but large numbers of physicians make regular use of empirical research in selecting their diagnoses and treatment regimes without participating in its production. Their training may enable them to be intelligent consumers of medical research, but this does not mean they have the skills or the motivation to conduct it on their own.

Similarly, our hope for a more robust research culture in forensic science would not turn every forensic scientist into a scientific researcher. Some practicing forensic scientists would no doubt participate in conducting research and, as we shall argue below, there ought to be greater incentives in place to create a larger pool of "two-hat" forensic practitioners—individuals who

---

39. On the ways that forensic scientists may feel "role ambiguity" that makes them uncomfortable with the idea of being research subjects, see Simon A. Cole, *Comment on 'Scientific Validation of Fingerprint Evidence Under* Daubert', 7 LAW, PROBABILITY & RISK 119, 122 (2008).

40. *See* Simon A. Cole, *Toward Evidence-Based Evidence: Supporting Forensic Knowledge Claims in the Post-*Daubert *Era*, 43 TULSA L. REV. 263 (2007). On evidence-based medicine, see, for example, DAVID L. SACKETT ET AL., EVIDENCE-BASED MEDICINE: HOW TO PRACTICE & TEACH EBM (1997); WILLIAM A. SILVERMAN, WHERE'S THE EVIDENCE? DEBATES IN MODERN MEDICINE (1998); STEFAN TIMMERMANS & MARC BERG, THE GOLD STANDARD: THE CHALLENGE OF EVIDENCE-BASED MEDICINE AND STANDARDIZATION IN HEALTH CARE (2003); Evidence-Based Medicine Working Group, *Evidence-Based Medicine—A New Approach to Teaching the Practice of Medicine*, 268 JAMA 2420 (1992); David L. Sackett et al., *Evidence-Based Medicine: What It Is and What It Isn't*, 312 BRIT. MED. J. 71 (1996); Stefan Timmermans & Aaron Mauck, *The Promise and Pitfalls of Evidence Based Medicine*, 24 HEALTH AFF. 18 (2005). For a classic reflection on the sometimes strained relationship between science, research, and the clinical practice of medicine, see LEWIS THOMAS, THE YOUNGEST SCIENCE: NOTES OF A MEDICINE-WATCHER (1983).

41. Many biomedical researchers have both PhDs and MDs, but there are also many research physicians with MDs alone.

are well trained in a forensic discipline and who have also received substantial formal training in research methodologies. More generally, rank-and-file forensic practitioners without special research-oriented training should be taught through both training and laboratory practice to have respect for research findings. Through appropriate hiring and training, forensic practitioners can understand the value of a research culture and apply research findings in practice. But just as a novice ought not to walk into a forensic science laboratory and begin analyzing casework, neither should forensic practitioners—even those with a bachelor's degree in a scientific discipline and a master's degree in forensic science—be expected, or even necessarily encouraged, to develop or execute a research program on their own. Many practitioners can certainly assist with research—not only by being test subjects, but also by helping to generate research agendas regarding the questions that would help them do their job more effectively, and by partnering with statisticians, psychologists, computer scientists, physical scientists, and research-oriented forensic scientists. But to reiterate: A research culture in forensic science would not turn most practicing forensic scientists into researchers.

What, then, is a research culture? We cannot succeed in providing a robust and complete definition of a research culture, nor shall we attempt to do so. But we can usefully describe core constellations of values that are necessary pieces of a well-functioning research culture in any discipline. We believe these core values are empiricism, transparency, and an ongoing critical perspective; we elaborate on each below.

A. Empiricism

A research culture should have a deep and fundamental respect for the ideal of empirical support. Claims, both about a field and about particulars, should be expected as a matter of course to be data-driven. Moreover, thoughtful attention should focus on the degree to which the body of available data supports any given claim, and on the relationship between research results, the claim made, and the degree of confidence expressed. Hunches—or claims based on anecdote or personal experience—ought not have the same status as knowledge justified by a substantial body of rigorously produced data. Research that is deeply methodologically flawed should be given no credence. Moreover, research that is methodologically sound should not be touted as offering support for propositions that extend beyond the reach of the research design. In short, the extent of sound empirical support for claims should guide practices in the laboratory, conclusions in reports, and testimony in the courtroom.

B. Transparency

A research culture maximizes transparency, both in the production of knowledge and in internal practices and procedures. Researchers should be encouraged to make data sets available to other researchers, both to share the particular basis for their own claims and to encourage further research. To the extent feasible, laboratories should assist in the production of data sets that can help address key research inquiries, and laboratory personnel should be willing to participate in research projects both as collaborators and as test subjects. To be sure, laboratories may need to delimit access for practical or legal reasons, and laboratory personnel may need to participate as research subjects only to the degree it does not interfere with ongoing operations. But access to data and to examiners as test subjects ought not to depend on being a practicing forensic scientist (as opposed to a researcher from another discipline), nor should it require giving a laboratory veto power or control over publication or dissemination of the results. More generally, information about ordinary laboratory practices, procedures, and protocols should be publicly available.

In addition, errors should be recognized as an inevitable part of any human enterprise. Errors should be acknowledged rather than swept under the carpet. Both the individual and the community should take the opportunity to learn from them. We do recognize that forensic laboratories and forensic examiners work within an adversarial legal system. Certainly the us-versus-them mentality that adversarialism generates can discourage disclosure beyond what is legally mandated. The dynamics of cross-examination, in which ordinary human limitations and innocent inconsistencies may be leveraged by opposing counsel into challenges to credibility, can exacerbate this tendency. These forces may combine to create significant pressures opposing transparency. While we do understand this tension—and in the final Part of this Article, we offer some suggestions for managing it[42]—we reiterate that transparency is a critical value of a functioning research culture.

C. Ongoing Critical Perspective

Claims of knowledge should be taken as provisional and subject to revision in the face of new information. Dogma should be resisted. Research is not one thing, or one study, or once done, never reexamined. Research is an ongoing, incremental process. Research problems should be approached with an open mind. While it is certainly appropriate to have a hypothesis, or preliminary

42. *See, infra*, Part IV.

expectation, about what any given research study will show, investigators should follow the data whether or not it supports their original hypothesis, and whether or not it legitimates current practices. Research projects should be designed according to the norms of relevant academic fields. They should not be designed defensively, to produce, or to increase the chances of producing, a particular outcome.[43] Publication and peer review should occur as a matter of course, and a commitment to publication should not depend on the results. At the same time, we must recognize that the questions that scientific research attempts to answer and the questions that must be answered in a courtroom during a trial are very different. Science is a moving target; answers are always provisional and can be updated as research produces new information or challenges accepted findings. But in a trial, the judge or jury must make pragmatic use of the best available answers to scientific questions at that given moment in time. As a result, the legal system may quite legitimately accept evidence, even scientific evidence, that is good enough rather than perfect.[44] Waiting for the next study, or postponing a decision, is typically not an option. But these determinations, while decisive in a particular case, should remain epistemically provisional, subject to critical inquiry and revision in a future case if the research warrants it.

III. THE PRESENT LACK OF A RESEARCH CULTURE IN FORENSIC SCIENCE

A growing number of individuals within the pattern identification disciplines and other forensic fields do fundamentally embrace the values associated with a research culture. Nonetheless, at present, these values are not sufficiently widespread within the pattern identification communities. In this Part, we provide a variety of examples that illustrate the ways in which a research culture is still weak or absent in these disciplines.

---

43. For an example of a recent research study that was criticized along these lines, see Lisa J. Hall & Emma Player, *Will the Introduction of an Emotional Context Affect Fingerprint Analysis and Decision-Making?*, 181 FORENSIC SCI. INT'L 36 (2008). For the criticisms, see Itiel E. Dror, *On Proper Research and Understanding of the Interplay Between Bias and Decision Outcomes*, 191 FORENSIC SCI. INT'L e17 (2009); Michael J. Saks, *Concerning L.J. Hall, E. Player, "Will the Introduction of an Emotional Context Affect Fingerprint Analysis and Decision-Making?"*, 191 FORENSIC SCI. INT'L e19 (2009). For the study authors' response to these criticisms, see Lisa J. Hall & Emma Player, *The Value of Practitioner Research in the Field of Fingerprint Analysis*, 191 FORENSIC SCI. INT'L e15 (2009) (responding to Dror's criticism); Lisa J. Hall & Emma Player, *The Value of Practitioner Research in the Field of Fingerprint Analysis*, 191 FORENSIC SCI. INT'L e21 (2009) (responding to Saks's criticism).

44. *See, e.g.*, Jennifer L. Mnookin, *Of Black Boxes, Instruments, and Experts: Testing the Validity of Forensic Science*, 5 EPISTEME 343 (2008); Dale A. Nance, *Reliability and the Admissibility of Experts*, 34 SETON HALL L. REV. 191 (2003).

In court, forensic analysts asked about the bases for their claims frequently refer to experience and training rather than providing any systematic data. Experience is a legitimate basis for certain kinds of knowledge, but it is deeply problematic for experience alone to be the basis for sweeping claims like individualization.[45] Moreover, without robust feedback mechanisms to detect and provide information about any possible mistake, experience cannot be a sound warrant for reaching valid conclusions.[46] If, for example, a document examiner generally has no independent knowledge of whether or not her conclusions in any given case are actually correct, how can she learn from her experience? If she never knows when or if she makes an error, how can she adjust her practices to increase accuracy? At present, the efforts to create these kinds of feedback mechanisms are minimal. For example, no laboratory of which we are aware regularly conducts blind proficiency tests that are given in the stream of casework in a pattern or impression discipline, or, for that matter, in any other forensic discipline. Airport security staff, by contrast, are frequently

45. Individualization is the assertion that an item can be identified to a unique, specific source—that a print can be identified to a particular finger, to the exclusion of every other finger in the universe, or that a handwriting specimen can be identified as belonging to one and only one particular author out of the entire human population. Because no individual examiner can ever examine every possible specimen in the universe, experience alone cannot justify a claim of individualization, assuming that the potential population of the source is substantial. *See, e.g.*, KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 7.3.2(c)(2), at 324 ("The least useful measures of errors are self-congratulatory statements by the practitioners of a technique. In the absence of systematic, unbiased efforts to root out errors, these estimates amount to little more than reports of the 'I don't remember being proved wrong' variety."). Whether individualization might ever be a plausible claim is a far more difficult question and one upon which the authors of this Article do not all agree. *See, e.g.*, Simon A. Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 LAW, PROBABILITY & RISK 233 (2009); Kaye, *supra* note 27; Jonathan J. Koehler & Michael J. Saks, *Individualization Claims in Forensic Science: Still Unwarranted*, 75 BROOK. L. REV. 1187 (2010); Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199 (2008). But we do agree that experience examining latent prints—even extensive experience looking at tens of thousands of prints—does not provide an adequate warrant for the assertion of individualization.

46. KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 10.3.3 ("Numerous studies have found that without quick and accurate feedback on correct and incorrect judgments, experience does not produce expertise and experts routinely overestimate their skills. . . . Casework in forensic handwriting analysis, latent fingerprint identification, toolmark identification, and other patterns and impression evidence comparisons rarely involve . . . feedback based on ground truth. The argument that the judgments of these analysts are valid merely because the practitioners have had specialized training or ample experience therefore is unimpressive."); *see* D. Michael Risinger & Michael J. Saks, *Science and Nonscience in the Courts:* Daubert *Meets Handwriting Identification Expertise*, 82 IOWA L. REV. 21, 33–34 (1996); Mnookin, *supra* note 44. Even outside forensic science, other disciplines vary with regard to the extent of feedback provided by experience. Physicians, for example, get more feedback than forensic examiners via patient outcomes, but this is a noisy signal—patients sometimes recover despite care rather than because of it, and even effective therapies may be ineffective in a given instance either due to bad luck or confounding issues. Mechanics, for example, have better access to feedback than either forensic scientists or doctors: Automobiles are not self-healing and their mechanisms are less complex than bodies.

tested covertly in a variety of ways as part of their ordinary workstream.[47] In one scheme, electronic images of dangerous materials are superimposed onto actual passengers' carryons. Other testing, conducted independently by several agencies, includes no-notice testing of inert bomb parts, weapons, and other prohibited materials. The TSA explains on its website, "Covert testing is a critical element of the aviation security system. It measures effectiveness, identifies vulnerabilities, constantly adapts to challenge officers while incorporating intelligence in a useable way. Simply put, without adopting difficult, covert testing, the aviation security system would not be as effective as it is."[48] Would forensic science not also benefit from covert testing? Another potentially beneficial technique for assessing strengths and vulnerabilities is randomly selected case audits to seek out mistakes or assess the quality of analyses conducted. Some laboratories do carry out such audits, but neither standard practice nor accreditation requirements insist upon it. Institutionalizing procedures like these would serve to check the quality and effectiveness of examiners' experience and would provide critical information about accuracy.[49]

47. The variety of tests employed are briefly described on the website of the Transportation Security Administration. *Covert Testing: Security Screening*, TRANSPORTATION SECURITY ADMINISTRATION, http://www.tsa.gov/what_we_do/screening/covert_testing.shtm [hereinafter TSA] (last visited Feb. 11, 2011); *see also* Dror, *supra* note 25, at 103.

48. TSA, *supra* note 47.

49. There are, to be sure, some proficiency tests currently in use. But they are not conducted blindly, nor are they necessarily performed by individual examiners working alone, without collaboration or assistance from colleagues. Nor, for the most part, does their difficulty level mirror actual casework. On the problems with the current proficiency tests in use in the pattern identification field and the potential for using proficiency tests as a method for assessing accuracy, see Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985 (2005); Lyn Haber & Ralph Norman Haber, *Error Rates for Human Fingerprint Examiners*, *in* AUTOMATIC FINGERPRINT RECOGNITION SYSTEMS 339 (Nalini Ratha & Ruud Bolle eds., 2004); Lyn Haber & Ralph Norman Haber, *Scientific Validation of Fingerprint Evidence Under* Daubert, 7 LAW, PROBABILITY & RISK, 87 (2008); Jonathan J. Koehler, *Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter*, 59 HASTINGS L.J. 1077 (2008); Mnookin, *Confessions*, *supra* note 35; D. Michael Risinger et al., *The* Daubert/Kumho *Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CALIF. L. REV. 1 (2002); Jonathan J. Koehler, Proficiency Tests to Estimate Error Rates in the Forensic Sciences 1–5 (Sept. 19, 2010) (unpublished manuscript) (on file with Northwestern University School of Law).

As for auditing, although the American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLD-LAB) does audit some cases as part of its accreditation process, this review takes place only once every five years, and cases reviewed are not selected at random. For a recent instance in which an accredited laboratory had significant problems in its blood analysis not found through the accreditation process, see Locke & Neff, *supra* note 5. On ASCLD-LAB Accreditation Requirements, see ASCLD/LAB-INTERNATIONAL, ASCLD/LAB, INC., INTERNATIONAL ACCREDITATION PROGRAM: PROGRAM OVERVIEW (2006).

Many forensic scientists, as well as many judges, are too willing to infer scientific validity from the fact of longstanding use.[50] It is true that some of these forensic techniques have been in use for a substantial period.[51] It is also true that the number of proven errors for some of these techniques is small relative to the frequency of use (though certainly greater than zero).[52] And it is true that those pieces of information, combined, provide a degree of support for the claim that latent fingerprint identification, for example, likely has a substantial degree of accuracy (though this evidence obviously does not permit quantification of the precise degree of accuracy). Furthermore, whenever a pattern analyst matches an exemplar to a source, and highly probative, independent evidence of guilt subsequently emerges (or already existed but was unknown to the examiner), this corroborating information provides some indication that the identification technique works, notwithstanding that case information alone can never provide absolute assurance about ground truth. Moreover, if these techniques were being widely used but misidentifying sources

50. *See* United States v. Llera Plaza, Nos. CR. 98-362-10, CR. 98-362-11, CR. 98-362-12 (E.D. Pa. Jan. 7, 2002) ("[T]he ACE-V process and the experts' conclusions have been tested empirically over a period of 100 years . . . ."); United States v. Havvard, 117 F. Supp. 2d 848, 854 (S.D. Ind. 2000) ("[T]he methods of latent print identification can be and have been tested. They have been tested for roughly 100 years. They have been tested in adversarial proceedings with the highest possible stakes—liberty and sometimes life."); Transcript of Trial, Day Three at 114–15, United States v. Mitchell, No 96-407 (E.D. Pa. July 9, 1999) ("[E]mpirical studies is when you roll up your sleeves, you do observational analysis. The idea of taking prints, comparing them to other prints to seeing how often things are similar or dissimilar, is empirical studies. The 100 years of fingerprint employment has been empirical studies.") (testimony of Bruce Budowle); David L. Grieve, *Simon Says*, 51 J. FORENSIC IDENTIFICATION 85, 95 (2001) ("The testability of fingerprint individuality has been conducted for nearly a century, perhaps not in one grand empirical study that captivated the [*Daubert*] defense, but in the countless smaller studies performed daily in all parts of the globe."); WillIam F. Leo, *Fingerprint Identification: Objective Science or Subjective Opinion?*, 17 PRINT 1, 2 (2001) ("A fingerprint examiner's knowledge and ability can be and is tested, is documented and can be verified, and is evaluated by the courts and juries every time the examiner takes the witness stand."); *60 Minutes*: *Fingerprints* (CBS television broadcast Jan. 5, 2003) ("We're winning 41 times out of 41 [admissibility] challenges. I think that says something. We have 100 years of experience; let's make sure that that's clearly out there. And if it wasn't reliable, this certainly would have been discovered many, many years ago."). *But see* Bruce Budowle et al., *A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement,* 54 J. FORENSIC SCI. 798, 799 (2009) ("[F]or many years the forensic science community has pointed to successful admissibility of its science findings, and the opportunity to cross examine expert witnesses, as support of a technique's 'general acceptance' and 'reliability'. . . . [P]hilosophically we do not advocate successful admissibility as demonstrating good science.").

51. For an example of the early history of fingerprint evidence, see generally COLE, *supra* note 35; Mnookin, *Fingerprint Evidence*, *supra* note 35. For the history of handwriting identification evidence, see generally Jennifer L. Mnookin, *Scripting Expertise*: *The History of Handwriting Identification Evidence and the Judicial Construction of Reliability*, 87 VA. L. REV. 1723 (2001) [hereinafter Mnookin, *Scripting Expertise*]; Risinger et al., *supra* note 35; D. Michael Risinger, Mark Denbeaux & Michael J. Saks, *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise"*, 137 U. PA. L. REV. 731 (1989).

52. Cole, *supra* note 49.

at an extremely high rate, one might expect that in some of these cases, powerful contrary evidence supporting innocence would emerge and throw the identification technique's general accuracy into doubt.[53]

While we do therefore grant that this longstanding use establishes something, it establishes less than its advocates suggest. First, the very fact that many kinds of pattern evidence are believed to be especially powerful and persuasive proof makes inferring validity from its success dangerous. If a fingerprint error leads to a misidentification, might the identified individual nonetheless be convicted, or even plead guilty to avoid a stronger sanction at trial, in the face of evidence that seems virtually indisputable? Convictions, therefore, do not necessarily establish the accuracy of the evidence undergirding them. To argue otherwise is a form of rhetorical bootstrapping. Further support of this point is provided by numerous failures to uncover errors until well after conviction and sometimes only through highly fortuitous circumstances.[54] Moreover, the growth of searchable databases with millions of latent fingerprints may create significant new dangers because a large database increases the chances of finding prints from different sources with a high degree of coincidental similarity.[55] In addition, defense challenges to fingerprint evidence, firearms comparison, and other pattern evidence, have been, until recently, very unusual;[56] as a result, these techniques have operated in court almost as if they were self-proving.[57]

The key point is that longstanding use leads some forensic scientists (and many judges) to treat questions of scientific and systematic validation as moot, or at a minimum, not terribly important.[58] A research culture would care

53. We make this last point with caution, because strong evidence can likely only be beaten by equally strong evidence. If, for example, fingerprints are widely seen as dispositive, the emergence of other evidence strongly suggesting innocence may be brushed aside as erroneous in the face of the fingerprint evidence.

54. Cole, *supra* note 49, at 1020–23.

55. Itiel E. Dror & Jennifer L. Mnookin, *The Use of Technology in Human Expert Domains: Challenges and Risks Arising From the Use of Automated Fingerprint Identification Systems in Forensic Science*, 9 LAW, PROBABILITY & RISK 47, 58 (2010). Databases also play a role in firearms comparison, though the scale of images in the database is significantly smaller than the largest automated fingerprint identification systems.

56. To be sure, in some cases defense counsel may consult with defense experts in fingerprint identification but elect not to present any defense challenge. While we have no data on the frequency of such consultations, our point is that the testimony has typically been presented to the factfinder unchallenged. *See generally* COLE, *supra* note 35; Mnookin, *Fingerprint Evidence*, *supra* note 35.

57. In the early history of handwriting cases, and at present, in civil disputes, document examination has tended to have competing experts on both sides. But this has not generally extended to criminal disputes, especially in modern times. *See* Mnookin, *Scripting Expertise*, *supra* note 51, at 1730; Risinger et al., *supra* note 35; Risinger & Saks, *supra* note 46.

58. *See, e.g.*, United States v. Havvard, 117 F. Supp. 2d 848 (S.D. Ind. 2000), *aff'd*, 260 F.3d 597 (7th Cir. 2001); United States v. Crisp, 324 F.3d 261 (4th Cir. 2003). For criticism of this approach as unfaithful to *Daubert*'s call for scientific validation, see KAYE, BERNSTEIN & MNOOKIN, *supra* note 3,

about, and be willing to invest in, rigorous empirical validation even of those matters widely thought to be obvious by practicing forensic scientists.

In addition, a research culture would realize that casework is not research. To be sure, researchers may introduce research questions into the stream of what looks to an analyst like ordinary casework. Covert research of this sort can provide some of the most ecologically valid data about actual practices. Research could also entail examining casework in a structured manner. But an analyst engaged in ordinary casework is not herself conducting research.[59] Casework may suggest research problems worth exploring. It may lead to hypotheses worth developing. Unusual case findings may be worth discussing at professional meetings or publishing as food for thought. Indeed, the International Association of Identification (IAI) routinely publishes such materials in its journal, and they may provide useful platforms for discussion and expand the experiential basis available to practitioners. But case findings ought not to be mistaken for structured research or empirical data that goes beyond the anecdotal, whether or not such findings are published.[60] Unlike planned research, casework does not permit the development of careful controls, defined independent variables, or structured and directed focus. Also, and critically, in casework, ground truth is not known and cannot simply be inferred by a conviction, a confession, or the consensus judgment of experts.

However, we do not mean to set up an unrealistically idealized vision of real research. Legitimate research can vary in its degree of formality and ecological validity. Often, very good research necessarily simplifies some aspects of the real world to focus attention on the matter at issue and to limit potentially confounding variables. Good research can, and usually does, involve both hard questions of design and imperfect compromises. But research does, and must, involve explicit study design. And research reports and publications, comporting with the research culture value of transparency, must be as explicit as is feasible about the nature of the study design.[61]

---

§ 7.3.2(a)(4) (arguing that adversarial testing is not scientific testing); *see also* Mnookin, *supra* note 4, at 36–37.

59. Simon A. Cole, *"Implicit Testing": Can Casework Validate Forensic Techniques?*, 46 JURIMETRICS J. 117 (2006).

60. For examples of the publication of such case studies from fingerprint identification, see Michael H. Kershaw, *Laterally Reversed*, 50 J. FORENSIC IDENTIFICATION 138 (2000); Robert D. Reneau, *Unusual Latent Print Examinations*, 53 J. FORENSIC IDENTIFICATION 531 (2003); Dana Shinozuka, *Fingerprints on a Banana Leaf*, 50 J. FORENSIC IDENTIFICATION 441 (2000).

61. Transparency does of course have its limits. Among other reasons, confidentiality concerns and maintaining the integrity of the project—which may mean, for example, that examiners do not necessarily know when they are being studied—may require a degree of secrecy. For discussions of the importance of study design, see, for example, KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 12.5 (discussing the importance of the design of studies); HANS ZEISEL & DAVID H. KAYE, PROVE IT WITH

Forensic analysts have often failed to recognize the limits of what conclusions are actually warranted by a given research result. Research is sometimes used to support conclusions that the data in question simply do not establish. For example, in fingerprint analysis, evidence that supports wide variation in human friction ridge detail is frequently offered to support the examiner's ability to match unknown prints to a source.[62] While the assertion that every fingerprint is different is an inductive claim that cannot definitively be proven empirically (because it is impossible to look at every fingerprint that has ever existed or will exist), the available empirical evidence does support the claims that a high degree of variation in human friction ridge detail exists and that an individual's friction ridges persist to a substantial degree over a lifetime.[63] Most of us would even be willing to infer, based on what is known, that every human being has prints observably distinguishable from those of every other at some "scale of detection."[64]

But this claim of variability of rolled or digitized fingerprints[65] does not establish that fingerprint examiners can therefore individualize prints recovered from crime scenes to a particular source or even that the techniques of fingerprint comparison necessarily "work." The right question is not whether

FIGURES: EMPIRICAL METHODS IN LAW AND LITIGATION 11 (1997) (discussing compromises such as "half-a-loaf" experiments to design workable studies); David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 83 (Federal Judicial Ctr. ed., 2d ed. 2000).

62. For a variety of examples of this rhetorical move, see those discussed in Cole, *supra* note 45, at 235–40. For a recent example of an analysis that makes use of this argument, see, for example, Peter E. Peterson et al., *Latent Prints: A Perspective on the State of the Science*, 11 FORENSIC SCI. COMM. (2009), *available at* http://www2.fbi.gov/hq/lab/fsc/backissu/oct2009/review/2009_10_review01.htm. This Article does recognize in passing that latent examiners "do not compare friction ridge skin directly" but rather examine two-dimensional representations that may introduce additional interpretive concerns. *Id.* But while numerous citations are offered in support of the premises of persistence and individuality of friction ridge skin, the only citation offered to support the claim that latent impressions "translate reliably as a true and accurate representation of what appears on the friction ridge skin" is an untitled FBI laboratory manuscript listed as "in preparation." *Id.*

63. *See, e.g.*, Christopher Champod & Pierre A. Margot, *Computer Assisted Analysis of Minutiae Occurrences on Fingerprints*, *in* PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON FINGERPRINT DETECTION AND IDENTIFICATION 305–18 (Joseph Almog & Eliot Springer eds., 1996); Anil K. Jain et al., *On the Similarity of Identical Twin Fingerprints*, 35 PATTERN RECOGNITION 2653 (2002); Neumann et al., *supra* note 32, at 54–64; Nicole M. Egli, Interpretation of Partial Fingermarks Using an Automated Fingerprint Identification System (Mar. 10, 2009) (unpublished Ph.D. dissertation, University of Lausanne) (on file with author); PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON FINGERPRINT DETECTION AND IDENTIFICATION: JUNE 26–30, 1995, NE'URIM, ISRAEL (Joseph Almog & Eliot Springer eds., 1996).

64. *See* KEITH INMAN & NORAH RUDIN, PRINCIPLES AND PRACTICE OF CRIMINALISTICS: THE PROFESSION OF FORENSIC SCIENCE 128 (2001).

65. A rolled fingerprint is an impression made with ink—usually black ink on white paper—where the individual rolls his inked finger to create a visible impression. Digitized fingerprints involve scanning the friction ridge impression electronically rather than using ink.

all fingerprints actually differ from each other, but rather what conclusions the methods of fingerprint comparison permit, and in what circumstances. Even if every set of ten prints is different from every other, two specific portions of two prints from different individuals might be extraordinarily similar to one another. And even if every area of friction ridge skin is different from every other individual's inked or scanned print, that does not answer whether two such prints from different sources might share enough similarity that an examiner, even if competently using the techniques of the field, might nonetheless mistakenly attribute them to the same unique source.

Moreover, latent print analysis involves difficulties often not present in the analysis of ten prints: Latent images are frequently smaller in surface area than the full print, they are possibly distorted, and they often contain artifacts resulting from the processes necessary to make a latent print visible.[66] So, the right question is whether, competently using the tools and techniques of latent fingerprint identification, two impressions from two different sources might ever be mistaken as coming from the same source (or, conversely, whether two impressions from the same finger might erroneously be said to come from different sources). Whether the actual ridge patterns on the two fingers in question are or are not "truly" the same is not the critical question. These are significantly different inquiries. The point is to recognize that the claim that friction ridge patterns are highly variable might be a necessary precondition for fingerprint identification, but it does not establish fingerprint analysts' ability to make a match. To suggest otherwise reflects a failure to think carefully and critically about the relationship between an empirical warrant and the claim that is being made.

Numerous examples within the forensic sciences reveal dogma or ideology trumping academic inquiry. For example, in 2001, two forensic science researchers, one of whom was a trained and qualified fingerprint examiner in Switzerland, published a commentary on fingerprint identification. In it, they called for abandoning "absolute conclusions."[67] The authors recognized the inherently probabilistic nature of fingerprint evidence; they allowed that the key question was not the uniqueness of friction ridge skin but rather the analyst's ability to recognize sufficient information from very limited information; and they advocated replacing experience-and-tradition-based approaches with more transparent and empirically justified practices.[68] At least one commentator

66. To be sure, not all fingerprint comparisons involve latent prints. Sometimes prints found in crime scenes are patent prints—left in ink, blood, or otherwise visible without dusting or processing.

67. Christophe Champod & Ian W. Evett, *A Probabilistic Approach to Fingerprint Evidence*, 51 J. FORENSIC IDENTIFICATION 101 (2001).

68. *Id.*

responded in print with explicit hostility at the notion that interloping statisticians would dare upset the apple cart.[69] This angry critic wrote:

> Once again, identification science is under attack, this time from a shotgun blast by statisticians. They come not to bury fingerprints but to praise it. But as with Shakespeare's Mark Antony, they actually come to incite a riot. Although their main point is relatively simple, it is mired deeply in rhetoric. One might describe it as opaque rather than transparent.[70]

The author later asserts:

> This commentary is indeed a vicious attack and any identification expert who does not see it as such has not read it closely enough. Surely the authors cannot expect that this will cause the scales to fall from the eyes of examiners everywhere and that the errors of the last hundred years will be revealed at last . . . . What then can be their motive for putting this forward at this time?[71]

He finally adds:

> As with most propaganda, it is masked, although not particularly well in this instance . . . . Although this article may be intended to demonstrate that identification specialists do not know enough about statistics, what it has clearly demonstrated is that statisticians do not know enough about identification.[72]

It is not clear which is more worthy of note: the vitriol and sarcasm of the response, or the fact that the journal published it notwithstanding this tone.

More recently, in the face of evidence presented in another article that contextual information may bias the decisions of fingerprint analysts,[73] one commentator responded with the following statements in a letter to the editor:

> *[A]ny* fingerprint examiner who comes to a decision on identification and is swayed either way in that decisionmaking process under the influence of stories and gory images is either totally incapable of performing the noble tasks expected of him/her or is so immature that he/she should seek employment at Disneyland. . . . And I do find it rather unsavoury that those within our own ranks, who ought to know better and are aware just how reliable the fingerprint system is, continue to provide fuel for

69. *See* Steve McKasson, *I Think Therefore I Probably Am*, 51 J. FORENSIC IDENTIFICATION 217 (2001).
70. *Id.* at 217 (citation omitted).
71. *Id.* at 221.
72. *Id.*
73. The study under discussion was Dror & Charlton, *supra* note 32.

> those within the media and Press who seem to relish attacking what is the most valuable tool in the investigating officer's armoury.[74]

Rather than discuss the merits of the research, the letter writer attacks those test subjects who showed themselves to be susceptible to biasing information as incompetent or immature. Given that psychological research shows that all humans are potentially susceptible to the effects of biasing information, this letter writer essentially proposed that fingerprint examiners might best make a mass exodus to Disneyland.

To be sure, these examples of blustery responses to unwelcome points of view obviously do not represent the views and attitudes of all forensic practitioners. But neither response provoked any apparent public outrage from the forensic science community. Not a single follow-up letter was published criticizing these authors for their sputtering and dogmatic responses to thoughtful research and analysis.

Admittedly, human endeavors are quite frequently dotted with examples of resistance to new theories that challenge the status quo.[75] Nonetheless, a sign of a mature discipline with a well-entrenched research culture is a willingness to engage respectfully with opposing viewpoints; it is a commitment to focusing on the merits of proposed theories, the adequacies of research methodologies, and the assessments of the data rather than resorting to inflated rhetoric or personal attacks. Forensic scientists have sometimes found it too easy to respond with a personal attack instead of—or layered on top of—substantive assessment of critics' arguments. Even one of the authors of this Article regrets portions of one of his early publications that now seem to him to have taken too derisive a tone toward some of the critics of forensic science (including, indeed, other authors of this Article).[76] In a research culture, participants should, ideally, learn from disagreements rather than fear them. We believe it is a significant step forward that those who have found themselves (literally) on opposite sides of the courtroom are now, by coauthoring this Article, not only willing to engage with one another, but are finding many shared views. But the development of a research culture in these areas still has a long way to go.

---

74. Martin Leadbetter, Letter to the Editor, FINGERPRINT WHORLD, Sept. 2007, at 231, 231.

75. Consider, for example, Thomas Kuhn's famous book, *The Structure of Scientific Revolutions*, and his arguments and examples detailing how "normal science" frequently proceeds even in the face of anomalous findings, and how most researchers in any given paradigm remain bound to it notwithstanding contradictory evidence. THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (1962). For further examples, see Bernard Barber, *Resistance by Scientists to Scientific Discovery*, 134 SCIENCE 596 (1961). Given these attitudes, even in areas with robust commitments to a research culture, it is not surprising to find similar dynamics in an area where the research culture commitment remains weak.

76. Glenn Langenburg, *Defense Against the Dark Arts*, CHESAPEAKE EXAMINER, Spring 2003, at 1, 5–6, 12.

When accused of being insufficiently research-based, or insufficiently linked to academia, practitioners in the pattern identification fields have sometimes responded by invoking the scientific foundations articulated by the pioneers of their fields. Whatever the qualifications of these early practitioners, in a healthy research culture, the scientific bona fides of a profession would be unlikely to depend on these pioneers of the distant past. For example, one published response to an article criticizing the lack of adequate scientific foundation in the forensic sciences emphasized the academic credentials of forensic pioneers like Calvin Goddard, J. Howard Mathews, and Sir Francis Galton.[77] Goddard (1891–1955) trained as a physician, spent his career partly in the military and substantially contributed to the establishment of forensic firearms comparison as a field. J. Howard Mathews (1881–1970) published a major firearms treatise in 1962,[78] ten years after retiring from an academic position in chemistry. Sir Francis Galton (1822–1911) was a significant scientific intellectual of the Victorian era, with interests as diverse as meteorology, eugenics, heredity, statistical analysis, and fingerprints.[79] Without belittling the significant intellectual contributions of these pioneers, a robust research culture should be continuous and current. Century-old work and the credentials of pioneers, however impressive, have little direct relevance to questions of present-day scientific legitimacy.[80]

Another major limitation of the current forensic science culture relates to several of the publication venues for the pattern identification field. Several of the most significant journals focused on publishing pattern identification research simply do not comport with broader norms of access, dissemination, or peer review typically associated with scientific publishing. For example, the *AFTE Journal*, a quarterly publication of the Association of Firearm and Toolmark Examiners, has published numerous articles on firearms identification.[81] WorldCat—the largest online catalog of library materials, which includes

77. CRIME LAB REPORT, FORENSIC PATTERN IDENTIFICATION: A HISTORY LESSON, AND SOME ADVICE, FOR SAKS AND FAIGMAN 4 (2009), *available at* http://www.crimelabreport.com/library/pdf/1-09.pdf.

78. 1 J. HOWARD MATHEWS, FIREARMS IDENTIFICATION (2d ed. 1973).

79. *See generally* MICHAEL BULMER, FRANCIS GALTON: PIONEER OF HEREDITY AND BIOMETRY (2003).

80. For similar observations concerning handwriting identification, see *id*. The handful of formal studies on the "black box" reliability of signature authentication has not changed the almost exclusive practical reliance on century-old sources. *Id*. at 773.

81. Many of the articles claimed by firearms analysts to validate their practices have been published in AFTE. *See, e.g.*, Ronald G. Nichols, *Defending the Scientific Foundation of the Firearms and Tool Mark Identification Discipline: Responding to Recent Challenges*, 52 J. FORENSIC SCI. 586 (2007). We take no position here on whether these publications and research bases provide an adequate foundation for the claims of firearms identification. Our point is that this journal is deemed by members of the community to

the holdings of 72,000 libraries worldwide, including virtually every university-based library in the United States—lists only eighteen libraries with a copy of this journal in their holdings.[82] Furthermore, the *AFTE Journal* does not appear to be indexed or included in any major indexing service anywhere.[83] The only available index to AFTE was created by an individual firearms examiner on his own initiative and was not continued past 2005.[84] Moreover, peer review of submissions to AFTE is not blind; the author and the reviewer are both aware of each other's identity.[85] In addition, the peer reviewers appear to come entirely from the editorial board, which consists entirely of AFTE members, and therefore includes no members from outside the toolmark and firearms practitioner community.[86] This journal therefore appears to have extremely limited dissemination beyond the members of AFTE itself; completely lacks integration with any of the voluminous networks for the production and exchange of scientific research information; and engages in peer review that is neither

---

be a critical publication venue. For a view critical of the research basis of firearms comparison claims, see Adina Schwartz, *A Systemic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification*, 6 COLUM. SCI. & TECH. L. REV. 1 (2005).

82. WorldCat lists a total of nineteen libraries in three separate file listings. However, the Library of Congress is listed twice. The American libraries that subscribe to AFTE are, in full: Cal State, Sacramento; Case Western Law School; George Mason University; George Washington University; Grambling State University; John Jay College of Criminal Justice; the Library of Congress; Mercyhurst College Hammermill Library; Ogeechee Technical College; Stetson University College of Law; Truman State University; and the Virginia Commonwealth University. To be sure, WorldCat's listings may to a certain degree understate access. First, despite its extensive inclusions, we recognize that some libraries are not in WorldCat. Second, we recognize that at some institutions, if an individual faculty member has a subscription, the library may elect not to pay for institutional access. However, this issue should have a potential effect on all journals associated with membership organizations, not simply forensic science journals; and by any standard, the number of research libraries subscribing to AFTE is remarkably small. Moreover, a faculty member with access significantly limits broader dissemination to those outside the individual faculty member's ambit.

83. This information comes from Ulrichsweb, an authoritative source of information on periodicals. *See generally* ULRICHSWEB—THE GLOBAL SOURCE FOR PERIODICALS, http://www.ulrichsweb.com/ulrichsweb/ (last visited Jan. 28, 2011).

84. This index can be downloaded at *AFTE Journal Keyword Index*, ASS'N OF FIREARM & TOOL MARK EXAMINERS (Oct. 24, 2005), http://www.afte.org/ExamResources/journalindex.htm. It appears to be the individual work of an Albuquerque Police Department firearms examiner.

85. See the description at *AFTE Peer Review Process*, ASS'N OF FIREARM & TOOL MARK EXAMINERS (Aug. 2009), http://www.afte.org/Journal/PeerReviewProcess.htm; *see also* Dominic J. Denio, *The History of the AFTE Journal, the Peer Review Process, and* Daubert *Issues*, AFTE J., Spring 2002, at 210, 210–14.

86. Indeed, AFTE membership is, for the most part, open only to practicing firearms and toolmark examiners (or those in training for the profession). For membership categories, see *AFTE Membership Information*, AFTE, http://www.afte.org/Membership/membership.htm (last visited Jan. 26, 2011). For a list of the editorial review panel, see *AFTE Journal Editorial Panel*, AFTE, http://www.afte.org/Journal/EditorsCommittee.htm (last visited Jan. 26, 2011).

blind nor draws upon an extensive network of researchers. None of this is compatible with an accessible, rigorous, transparent culture of research.[87]

The *Journal of Forensic Identification* (*JFI*), the journal of the International Association of Identification, suffers from similar limitations, though to a slightly lesser degree. WorldCat reports seventy-two libraries that contain print holdings of the journal and 123 that subscribe to the electronic version through ProQuest.[88] The *JFI* is included in a few major indexes, including ProQuest and SCOPUS. This quantum of accessibility may be adequate to permit an intrepid researcher to locate materials published within *JFI*. But it still fails to meet conventional standards of research access.[89] Indeed, the *JFI* is not even listed in the Web of Science, a large collection of more than ten thousand journals over a wide range of areas. Like the *AFTE Journal*, it is not analyzed in the databases assessing journal impact.[90] The *JFI* also gives its authors plaques

87. For a discussion of the nature of scientific peer review, see KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, § 7.3.2(b).

88. *JFI* does come in an electronic version, but it is available to libraries only with purchase of a large and expensive criminal justice periodicals package, rather than by itself. This also suggests that some of the electronic holders were not specifically choosing the *JFI* but received it along with whatever sources led them to the aggregate database. Whatever their motivation, access is access, and the subscribers to the larger database are providing access to those with access to that library. Some institutions subscribe to both the electronic and the print versions, so the total number of libraries providing access to the journal is slightly fewer than adding the two numbers would suggest. However, it appears that WorldCat likely understates electronic access, as not every library that lists with WorldCat lists every electronic holding they receive as part of a package. A call to ProQuest confirmed that *JFI* is not available for subscription alone but is a part of the Criminal Justice Periodicals Index. The ProQuest representative indicated that there are more than two hundred subscribers to this database but was unable to provide any more exact figures. Assuming that this number is accurate, it suggests either that some subscribers are not members of WorldCat, some subscribers are not listing their electronic access on WorldCat, or, as is most likely, a combination of both. Therefore WorldCat's numbers for electronic access need to be taken with a grain of salt. However, this electronic access subscription number for *JFI* can still be loosely compared with that of other journals. There is no reason to believe that libraries would be less likely to report this specific holding as opposed to other electronic holdings, so relative comparisons are likely meaningful, even if the specific number cannot be trusted.

89. It would be unfair to compare *JFI* to the major publishing venues of a broad scientific or social scientific discipline, as pattern identification is a subfield of forensic science. A more reasonable comparison might be, for example, *Social Studies of Science*, the journal published by an academic association (the Society for the Social Study of Science) with many fewer members than the IAI and associated with an extremely small academic subfield (sociology of science). This journal is listed on WorldCat as having 543 subscribers to the print version and 712 to the internet version. Like the *JFI* numbers, this electronic number likely understates access, probably even more substantially than *JFI*. *Social Studies of Science* is available as a package through its publisher, Sage. It is not clear that libraries providing access through Sage, or through the widely available JSTOR, would list such access on WorldCat, or that those libraries that provide both print and electronic access would include two separate listings in the catalog.

90. Impact ratings are an effort to evaluate how much scholarly "impact" specific journals or individual articles may have. Impact ratings, which focus on how often journal articles are cited, are imperfect proxies for journal influence and quality. Nonetheless, it is fair to conclude that an unrated journal has a low impact.

to mark the fact of publication. While this could be viewed as a nice gesture to recognize an author's efforts and to spur submissions, it is certainly not a practice widely seen in other disciplines, and it implicitly treats publication as an unusual accomplishment, rather than an expected consequence of engaging in research. *Fingerprint Whorld*, a quarterly United Kingdom–based journal that provides another important source of information to fingerprint examiners, is similarly difficult to acquire through libraries.

Several other publications include pattern-identification-related articles, most notably the American Academy of Forensic Sciences' *Journal of Forensic Sciences* (*JFS*);[91] and *Forensic Science International* (*FSI*) (published in Europe). These journals have a significantly greater degree of library dissemination and meet more of the typical indicia expected for research journals. They are widely indexed (including in SCOPUS, Pubmed, Medline, Web of Science, and numerous other locations), and they are included among the 7300 scientific journals that are assessed for impact by the ISI/Web of Knowledge. However, it is perhaps worth noting that none of the top fifty most cited articles in either *JFS* or *FSI* relates to pattern identification.[92] This does not discredit those articles in pattern identification that do appear in *JFS* and *FSI*. It illustrates, however, both that pattern identification disciplines make up only a small portion of the journals' overall focus and that none of the journals' most well-known and widely cited articles come from these fields.[93] While the *JFS* is both peer reviewed and adequately disseminated to a broad research and practitioner community, from the perspective of generating a robust research culture in the forensic sciences, one aspect of the AAFS policy is troubling: Presenting new research at the AAFS annual meeting obligates the presenter to give the *JFS* a right of first refusal (albeit unenforceable) on the relevant material.[94]

---

91. WorldCat has 919 listings for the print version and 301 for the electronic version of *JFS*. The same caveats about these numbers apply. Note also that the journal is listed multiple times, and these numbers derive from adding the various listings without cross-checking for possible duplicate listings. Note also that many print subscribers likely also have electronic access.

92. This was established by searching the ISI/Web of Knowledge by journal title and sorting by times cited. For a broader (but slightly dated) analysis of what topics in forensic science are highly cited, see Alan W. Jones, *Which Articles and Which Topics in the Forensic Sciences Are Highly Cited*, 45 SCI. & JUST. 175 (2005). In Jones's analysis, the topics garnering the most citations came from toxicology, criminalistics (almost entirely DNA-related), and pathology. *Id.* at 178–80. Whether these disparities are solely the result of population differences across different forensic specialties, or also reflect meaningful differences in the quantum of research engagement, cannot be determined without further study.

93. To be fair, pattern identification fields make up a relatively small portion of the total membership of the AAFS. However, this underscores the value of having a serious, well-disseminated journal focusing on these areas in particular.

94. *See Info for Authors*, AM. ACAD. OF FORENSIC SCIS., http://www.aafs.org/info-authors-0 (last visited Jan. 28, 2011) ("*JFS* reserves the right of first consideration for publication of any work accepted for presentation at an annual meeting of the AAFS, and authors must not submit their work elsewhere

While the *JFS* is a reputable journal, a researcher's publication options should not be restricted because of presentation to the forensic science community. Forensic science would benefit from broader dissemination and more frequent publication in high-impact journals that are not geared exclusively to the forensic sciences.[95]

While we firmly believe that an adequate research culture does not yet exist in the pattern and impression evidence disciplines, and is distressingly weak throughout many areas of forensic science, we are more interested in thinking constructively about how to remedy this situation than in pointing fingers and assessing blame. To prevent misunderstanding, it is worth making several points explicit. First, in our view, this lack of a research culture is not forensic scientists' fault. The two most significant causes are a dearth of funding and the fact that prosecutors, investigators, and the courts are the primary clients of forensic science. Until recently, very little federal grant money was available for non-DNA forensic science research.[96] This lack of funding, combined with the general paucity of resources in triage-driven, overworked laboratories, made research an exceedingly unlikely central priority. In addition, few practitioners had the background skills to develop substantial research programs even if the institutional climate had supported it.

Equally significant, even after *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[97] emphasized the need for judicial gatekeeping to assure the validity of expert evidence in court, most judges confronted with pattern identification evidence have continued to admit it without restriction.[98] If courts are not going to insist

---

for a period of six months following the annual meeting at which the work was presented. If a manuscript has not been accepted for publication, or is not under active consideration by *JFS*, at the end of the six-month period, the interest of *JFS* in the manuscript automatically terminates.").

95. Given this rule, it is not surprising that one recent study found that a majority of those papers presented at the AAFS that were later published in a peer-reviewed journal were published in the *JFS*. Silvia Tambuscio et al., *From Abstract to Publication: The Fate of Research Presented at an Annual Forensic Meeting*, 55 J. FORENSIC SCI. 1494, 1496 (2010). The same study also found that, in the annual meeting studied, only 16.4 percent of research presentations led to publication, a lower number than the vast majority of presentation-to-publication ratios that have been studied. *Id.* This unusually low publication ratio is yet another indicator of the lack of a robust research culture.

96. Max M. Houck, *A Vicious Cycle*, 1 FORENSIC SCI. POL'Y & MGMT. 123, 124 (2009).

97. 509 U.S. 579 (1993).

98. For discussion of these admissibility challenges, see FAIGMAN ET AL., *supra* note 3. Some recent cases evincing more skepticism, though generally still admitting the evidence, are discussed in KAYE, BERNSTEIN & MNOOKIN, *supra* note 3, and Mnookin, *supra* note 4, at 1212–13, 1241–65. For an interesting procedural order from one district court judge, see Procedural Order: Trace Evidence, No. 1:08-cr-10104-NG (D. Mass. Mar. 8, 2010), *available at* http://www.swgfast.org/Resources/100310-GertnerProceduralOrder.pdf (making clear that in the wake of the NAS Report, admissibility of such forensic science evidence "ought not to be presumed; that it has to be carefully examined in each case, and tested in the light of the NAS concerns, the concerns of *Daubert/Kumho* case law, and Rule 702 of the Federal Rules of Evidence," and describing pretrial procedures to govern any such challenges).

upon better evidence of validity, if they are instead going to continue to permit forensic scientists to reach extremely strong conclusions about their own abilities to make identifications, and if legal challenges remain both relatively rare and generally unsuccessful, then why should the forensic science community consider changing its practices? If an examiner is permitted, indeed expected, to express extremely high confidence about an individualization, what incentives exist to pursue research that would, at best, justify this confidence, and at worst, reveal hitherto unrecognized limitations? The judicial response to these identification techniques has therefore been a powerful force both enabling and preserving this status quo. If a few more brave judges had required additional evidence to support the claims being made and mandated a closer fit between claims made and the research supporting them, the forensic science community would have had an extremely strong incentive to develop and provide precisely this information.[99]

Moreover, most practicing forensic scientists in pattern and impression evidence, and in most other forensic disciplines as well, are not actually qualified to pursue the necessary research. Until recently, many laboratories did not necessarily require a college degree or any formal science training.[100] Even those with a BS in forensic science or some other scientific discipline have not typically received significant training in the development of research design. Experience may provide the basis for determining what questions to ask, but most pattern identification analysts, even with entirely noble intentions, would not be qualified to design or develop sophisticated research projects to answer those questions. We neither fault these practitioners for failing to do so, nor do

---

99. *See* Mnookin, *supra* note 4; Risinger & Saks, *supra* note 46, at 65–66; D. Michael Risinger, *Goodbye To All That or a Fool's Errand, by One of the Fools: How I Stopped Worrying About Court Responses to Handwriting Identification (and "Forensic Science" in General) and Learned to Love Misinterpretations of* Kumho Tire v. Carmichael, 43 TULSA L. REV. 447, 471–75 (2007). Some judges have evinced genuine concern about whether some pattern identification passes *Daubert*, and some have restricted the evidence (for example, by permitting descriptions of similarities but no conclusion regarding identity, or by prohibiting claims of absolute certainty about identity to the exclusion of all others) or occasionally excluded it. For examples of these approaches, see United States. v. Taylor, 663 F. Supp. 2d 1157 (D.N.M. 2009); United States v. Green, 405 F. Supp. 2d 104 (D. Mass. 2005); United States v. Llera Plaza, 179 F. Supp. 2d 492 (E.D. Pa. 2002), *vacated*, 188 F. Supp. 2d 549 (E.D. Pa. 2002); United States v. Hines, 55 F. Supp. 2d 62 (D. Mass. 1999); Maryland v. Rose, No. K06-0545 (Md. Cir. Ct. 2007). However, most admissibility challenges have resulted in the admission of the pattern evidence without restriction.

100. Peterson et al., *supra* note 62 (noting that while in the past "examiners were required to have, at a minimum, a high school diploma," many labs are increasing educational requirements). The Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) recommends that new entrants to the field have a minimum of a college degree from an accredited institution that included scientific coursework. *See* SCIENTIFIC WORKING GROUP ON FRICTION RIDGE ANALYSIS, STUDY & TECHNOLOGY, STANDARDS FOR MINIMUM QUALIFICATIONS AND TRAINING TO COMPETENCY FOR FRICTION RIDGE EXAMINER TRAINEES (2010), *available at* http://www.swgfast.org/documents/qualifications-competency/100310_Qualifications_Training_Competency_FR_1.0.pdf.

we expect them to become primarily focused on research themselves. We do, however, expect them to become more sophisticated in thinking about data and the legitimacy of inference. Expecting most practicing pattern analysts to become PhD-level researchers is not realistic, nor is it even a good idea. If, however, practitioners at all levels operated within a research culture, they would hone their critical thinking skills and regularly question what basis supports their claim to knowledge, both in an individual case and more broadly in a given discipline. Moreover, while all laboratory personnel need not, and indeed should not, be researchers themselves, it would not be unrealistic to require certain key personnel—perhaps the lead technical worker in a unit, and whoever is authorized to approve standard operating procedures—to have some minimum research qualification and experience.

IV. CREATING A RESEARCH CULTURE: SOME POSSIBLE STEPS TOWARD CHANGE

Culture is sticky. We fully recognize that cultural change does not come easily, and we do not mean to assume naively that the culture of pattern identification can be modified with ease. We do believe that the current controversies, the NAS Report, and its aftermath create the opportunity for both greater self-reflection and cultural change. We already see a number of positive developments and glimmers of future changes on the horizon.[101]

In this final Part of the Article, we wish to describe briefly a variety of steps that could help to create and institutionalize a research culture within the pattern identification sciences. Is every one of these necessary? Taken together, would they be sufficient? We are not certain of the answer to either of these

101. For a few examples of interesting developments, see Procedural Order: Trace Evidence, *supra* note 98 (a procedural order by a district court judge signaling a clear willingness to take the issues raised by the NAS Report seriously); IAI RESOLUTION, *supra* note 23 (reflecting a "change [in] the official position of the Association related to Friction Ridge Examinations based on advances in the science and scientific research" by no longer prohibiting fingerprint examiners from testifying in probabilistic language); the current NIST/NIJ working group on Human Factors in Friction Ridge Identification, which has brought together a broad range of perspectives and signals a welcome willingness of leaders of the fingerprint community to engage with academics ranging from statisticians to law professors; Cognitive Bias and Forensic Science Workshop: Northwestern University Law School, http://www.law.northwestern.edu/faculty/conferences/workshops/cognitivebias/ (last visited Jan. 14, 2011) (detailing an NSF-funded workshop at Northwestern bringing together cognitive psychologists not previously involved in forensic inquiries with forensic science practitioners); NAT'L SCI. & TECH. COUNCIL, COMM. ON SCI., SUBCOMM. ON FORENSIC SCI., http://www.forensicscience.gov (last visited Jan. 28, 2011) (describing the creation by the White House's Office of Science and Technology Policy of a Subcommittee on Forensic Science, "to assess the practical challenges of implementing recommendations in the 2009 National Research Council (NRC) report," and to advise the White House regarding how to achieve the report's goals).

questions, but we do believe that these suggestions would offer meaningful and constructive steps toward positive change.

Our suggestions also reflect an effort to be realistic about what is possible. For this reason, we are not calling for the courts to transform their approach to the admissibility of forensic science. Many (though not all) of us believe that this would be intellectually appropriate and, while potentially disruptive in the short run, could also have beneficial cultural effects in the medium term. If, for example, courts insisted on better error-rate information as a precondition for admissibility, the incentives for its production would dramatically increase. Given that the legal system is the major client for forensic science, the requirements courts impose will naturally, and perhaps inevitably, influence what quantum and what kinds of research are deemed necessary by the community itself. Indeed, to a significant degree, the current state of affairs is the direct product of the courts' nearly nonexistent gatekeeping for these forms of evidence. Had the courts applied *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[102] with an intensity in the forensic sciences similar to that seen in, say, the toxic torts arena, there is little doubt that the forensic science community would have become forceful advocates for whatever research seemed necessary to justify admissibility.[103] Instead, while some judges have engaged in a certain degree of hand wringing, few have actually insisted upon empirical data to support forensic examiners' claims. Unfortunately, given their responses to forensic science challenges over the past few years, the *Daubert* test's fuzziness and flexibility, and the limited appellate review that an 'abuse-of-discretion' standard provides, there is little reason to believe that the judiciary will become a force that spurs cultural transformation in the forensic sciences.

Most of us support the idea of creating an independent entity (such as the National Institute of Forensic Science (NIFS) recommended by the NAS) that supports and governs the forensic science community, including its research activities. We believe that the major reforms that we and others have called for would best be accomplished via a corresponding structural change and through the leadership and oversight that a new agency, if carefully conceived and implemented, could provide. But while the Senate Judiciary Committee is considering a legislative proposal that may create an independent entity within the National Institute of Justice to pursue some of what NIFS might have

102. 509 U.S. 579 (1993).

103. *But see* Joseph Sanders, *Applying* Daubert *Inconsistently?: Proof of Individual Causation in Toxic Tort and Forensic Cases*, 75 BROOK. L. REV. 1367 (2010).

accomplished,[104] the politics of creating a new federal agency seem unmanageable at the moment.

Given our pessimism regarding the likelihood that the courts will be major agents of change, or that a new agency will transform these fields, what, then, can and should be done to improve the research culture within these fields?

A. Increased Funding

One of the biggest obstacles to forensic science research has been the absence of specific federal funding to support it. The National Science Foundation (NSF) has at times funded forensic science research projects,[105] but the NSF focuses on fundamental, rather than applied research. Some of the necessary research within these fields may make important methodological and theoretical contributions to broader disciplines, such as probability theory, statistics, decision research, and cognitive psychology. These kinds of projects might be appropriate for NSF funding. But much of the research critical for the forensic sciences may not make a novel methodological or theoretical contribution to other academic fields. NSF does not traditionally fund these more applied forms of inquiry. Forensic science research (apart from DNA profiling) has not received significant funding through other sources either.[106] For example, until very recently, the National Institute of Justice (NIJ) funded very little foundational research in the pattern identification sciences.[107] This has begun to change in the last year, partly as a function of the NAS Report itself. In 2009, NIJ posted a solicitation for funding up to $10 million to applications proposing "Fundamental Research to Improve Understanding of the Accuracy,

104. *See* Int'l Ass'n for Identification, Preliminary Outline of Draft Forensic Reform Legislation (May 5, 2010), *available at* http://www.theiai.org/current_affairs/20100505_Draft_Outline_of_Forensic_Reform_Legislation.pdf; *see also* Letter From Joseph P. Bono, President, Am. Acad. of Forensic Scis., to The Honorable Patrick J. Leahy, Chairman, Senate Comm. on the Judiciary (June 14, 2010), *available at* http://www.aafs.org/sites/default/files/pdf/AAFSResponseToDraftOutline14June2010.pdf.

105. A search of the NSF database reveals at least six funded projects that are squarely connected to forensic science: John Beatty, *Dissertation Research: Taming the Hypervariable Witness: The Introduction, Contestation, and Regulation of Forensic DNA Evidence in the American Legal System*; Marcus Boccaccini & Daniel Murie, *Why Do Forensic Evaluators With Access to the Same Information Come to Different Conclusions When Retained by Opposing Sides in Legal Proceedings?*; Sarat Dass, *Statistical Methods for Fingerprint Image Analysis*; Jonathan Koehler, *Understanding and Improving Jurors' Use of Highly Diagnostic Statistical Evidence*; Jonathan J. Koehler, *Cognitive Bias and Forensic Science*; William C. Thompson, *Jurors' Evaluations of Forensic Science*.

106. *See, e.g.*, Houck, *supra* note 96, at 124 (showing paltry federal research funding for forensic science compared with other fields of science and engineering).

107. An earlier NIJ solicitation was withdrawn in the wake of an early *Daubert* challenge to fingerprint evidence. *See* United States v. Mitchell, 365 F.3d 215, 232 (3d Cir. 2004).

Reliability, and Measurement Validity of Forensic Science Disciplines."[108] Similar solicitations appeared in 2010.[109] In addition, the National Institute of Standards for Technology (NIST), the Department of Defense, the Department of Homeland Security, and the FBI have begun to provide some additional, albeit still limited, funding for pattern identification research.[110]

These numbers, while far better than nothing, are a drop in the bucket.[111] More funding—and stable and consistent forms of funding—is critical for a research culture to take root and flourish. These funding sources sorely need to be independent from law enforcement. Solicitations should be as broad in scope and as widely disseminated as possible to encourage greater involvement from discipline-based academic researchers from fields like physical science, psychology, statistics, and computer science. Funding could help attract creative, cutting-edge work from diverse researchers applying the methods and techniques of their fields. While forensic science has not typically been a domain of major inquiry for these disciplines, substantial funding will likely pique the interest of some academic researchers from a variety of disciplines.[112]

---

108. NAT'L INST. OF JUSTICE, U.S. DEP'T OF JUSTICE, SOLICITATION: FUNDAMENTAL RESEARCH TO IMPROVE UNDERSTANDING OF THE ACCURACY, RELIABILITY, AND MEASUREMENT VALIDITY OF FORENSIC SCIENCE DISCIPLINES (2009), *available at* http://www.ncjrs.gov/pdffiles1/nij/sl000878.pdf.

109. Several of the coauthors of this Article have applied for and/or received NIJ funding under these solicitations and others.

110. For example, see the listings at *Investigative Support & Forensics (ISF)*, TECHNICAL SUPPORT WORKING GROUP, http://www.tswg.gov/subgroups/isf/isf.html (last visited Jan. 26, 2011).

111. Houck, *supra* note 96, at 123–24.

112. We note that selection of funding recipients must also operate in accordance with the values of a research culture. Nonresearching practitioners should not substantively evaluate the research design merits of proposals, except in relation to practical concerns about which their experience produces expertise. Practitioners' views on what research questions are important, and why, can absolutely be considered, and if research proposals make unwarranted or naïve assumptions about how laboratories operate, that too is relevant to evaluation. But the academic merits of any given research design should be assessed by those with the research qualifications to evaluate them. A recent report revealed a failure of precisely these values at the National Institute of Justice, which suggests that it might be a problematic choice to spearhead the forensic reform effort. OFFICE OF THE INSPECTOR GEN., U.S. DEP'T OF JUSTICE, U.S. DEPARTMENT OF JUSTICE AUDIT OF THE NATIONAL INSTITUTE OF JUSTICE'S PRACTICES FOR AWARDING GRANTS AND CONTRACTS IN FISCAL YEARS 2005 THROUGH 2007, at xxiii–xxiv (2009). While we strongly advocate greater funding for fundamental forensic science research, we recognize that funding will only produce useful research and strengthen the research culture if it is administered and distributed in ways concordant with the values of a research culture. It is worth noting that both institutional capacity concerns and the need for independence from law enforcement pressures were reasons that the NAS Report strongly urged the creation of NIFS as an entirely new and independent agency. *See* NAS REPORT, *supra* note 8, at 20–21.

B. Improving Forensic Education to Enhance a Research Culture

In addition to encouraging greater participation from university-based researchers from a variety of fields, we strongly believe that forensic science would benefit from the emergence of a cohort of individuals with the skills and the background to operate both in the academic research community and in the world of practitioners. Currently, in the pattern identification field, the number of practicing analysts with a PhD in any discipline is quite small indeed. (This is in stark contrast to a number of other forensic fields, including DNA analysis and toxicology, in which a significant number of analysts hold PhDs).[113]

The majority of forensic practitioners in pattern identification need not—and should not—pursue PhDs. But if some relatively small fraction of practitioners were full citizens of both the world of research and the world of practice, it would offer enormously beneficial spillover effects. These practitioner-researcher hybrids could wear two hats by being true insiders in both communities. They would be valuable translators, mediators, and educators in both domains. They could both convey to fellow practitioners the need for a research-based approach and contribute to ensuring that research focuses on areas of genuine and important concern to practitioners.[114]

Given the significant value that would result from encouraging a small number of two-hat researcher-practitioners, the government ought to consider funding generous competitive grants for highly qualified pattern identification practitioners to pursue advanced graduate training in relevant disciplines, such as physical science, statistics, cognitive psychology, computer science, or at a research-focused forensic science program. These grants could, for example, pay half of an analyst's salary for a period of several years to allow the time and financial resources to pursue a PhD. A few highly competitive and well-funded grant opportunities of this kind would significantly contribute to the research culture of forensic science.

Another important step for creating and nurturing a research culture is the creation of research-based forensic science programs within academic institutions. While two-hat experts with a PhD in a substantive non-forensic field along with practical forensic experience can be key mediators between a

113. Admittedly (and perhaps ironically), our evidence for both of these claims is anecdotal and based on experience rather than the product of careful empirical study.

114. Of course, it is theoretically possible that they would be captured by one perspective or the other and either lose all touch with practical concerns or become highly credentialed spokespeople for the status quo. We think, however, that precisely because culture is sticky, significant exposure and integration into both domains will more likely produce individuals who, like those truly bilingual in two languages, can mediate, engage, and translate in both worlds.

research culture and forensic practice, research programs also have a place within forensic science departments. Academic forensic research programs will not generate a research culture in the forensic industry, but a small number of excellent research-oriented graduate programs in forensic science could help promulgate a research culture and could also produce valuable research. At present, most university-based forensic education is far more focused on training future practitioners than on training students to engage in fundamental research. We see doctoral-level training in forensic science as a supplement to, rather than a substitute for, this appropriate focus.

This is not a new idea. A few institutions, both past and present, have trained doctoral students to conduct significant and foundational research. For example, at the University of California, Berkeley, Paul Kirk, and later, John Thornton, supervised a number of doctoral dissertations on the quantitative and theoretical aspects of, among other topics, identification evaluation,[115] typewriting identification,[116] handwriting identification,[117] and fingerprint identification.[118] This was Kirk's deliberate attempt to help generate fundamental, theoretical research.[119] Research from other early forensic science programs, such as Michigan State University under Ralph Turner, also contributed significantly to the literature, even though such programs did not offer doctoral-level training.[120] Several strong European examples exist as well, both historically and at present.[121]

For the most part, however, there has been a disjunction between academic research and the forensic laboratory. Forensic laboratories in the United States

---

115. Charles R. Kingston, *Applications of Probability Theory in Criminalistics*, 60 J. AM. STAT. ASS'N 70 (1965); Q.Y. Kwan, Inference of Identity of Source (1977) (unpublished Ph.D. dissertation, University of California, Berkeley) (on file with author).

116. David Allan Crown, A Statistical Evaluation of Typewriting Individuality (June 14, 1969) (unpublished Ph.D. dissertation, University of California, Berkeley) (on file with author).

117. Edward Franklin Rhodes III, The Implications of Kinesthetic Factors in Forensic Handwriting Comparisons (1978) (unpublished Ph.D. dissertation, University of California, Berkeley) (on file with author).

118. Charles Richard Kingston, Probabilistic Analysis of Partial Fingerprint Patterns (1964) (unpublished Ph.D. dissertation, University of California, Berkeley) (on file with author). David Alan Stoney, A Quantitative Assessment of Fingerprint Individuality (Dec. 17, 1985) (unpublished Ph.D. dissertation, University of California, Berkeley) (on file with author).

119. *See* Paul L. Kirk, *The Ontogeny of Criminalistics*, 54 J. CRIM. L. CRIMINOLOGY & POLICE SCI. 235 (1963).

120. Ralph F. Turner, *Forensic Science Education—A Perspective*, *in* FORENSIC SCIENCE 1 (Geoffrey Davis ed., 1975).

121. Of particular note is the forensic science program at UNIL, in Lausanne, Switzerland. *See Forensic Science Department*, UNIV. OF LAUSANNE, http://www.unil.ch/esc/page10116.html (last visited Jan. 26, 2011).

were, from their outset, framed as arms of law enforcement and embedded within a different system of values from academia:

> Without a doubt, the laboratory, as it exists in the United States, is an appendage of a quasi-military operation of an enforcement agency. As in the military, the laboratory technician in the quasi-military operation is subordinate to the administration, which is usually not technically trained. The technician, therefore, does not have the freedom of decision nor the opportunity for research that would exist if he were a dedicated, well-trained scientist acting as a civilian in the proper framework.[122]

In some ways, the historical origin story of the forensic laboratory explains the divide between research values and forensic practice. The forensic laboratory, from the outset, was seen as bringing cutting-edge science to enhance older investigative methods (the "needle in the haystack" method of human intelligence and shoe leather).[123] The desirability of a laboratory—touted by the media as a new method of catching criminals[124]—led to the hasty but enthusiastic creation of new laboratories. This perhaps contributed to the inadequate delineation of roles between traditional investigators and scientific crime-fighters, and the extent of oversight of scientists by sworn officers. In the meantime, though eager for laboratories in principle, police departments did not always appreciate what they offered, nor did they understand how to make use of them in practice; at the extreme, they were an "incomprehensible"[125] novelty to the nonscientific police:

> Some of these [law enforcement] agencies which are so eager to have a laboratory have demonstrated to the author's satisfaction that they don't even know what a laboratory is for. Even worse, they have little or no conception of the proper use of a laboratory.[126]

---

122. C. Wilson, *Crime Detection Laboratories in the United States*, *in* FORENSIC SCIENCE: SCIENTIFIC INVESTIGATIONS IN CRIMINAL JUSTICE 96, 99–100 (J.L. Peterson ed., 1975).

123. COLIN WILSON & DAMON WILSON, WRITTEN IN BLOOD 18–19 (2003).

124. For an instance of *plus ça change plus c'est la même chose*, see Max M. Houck, *CSI: Reality*, SCI. AM., July 2006, at 85.

125. W. Fong, *Criminalistics and the Prosecutor*, *in* FORENSIC SCIENCE: SCIENTIFIC INVESTIGATION IN CRIMINAL JUSTICE, *supra* note 122, at 371.

126. Wilson, *supra* note 122, at 100; *see also* R. PERKINS, ELEMENTS OF POLICE SCIENCE 39–40 (1942) ("In general, American detectives do not place much weight upon the application of scientific principles to the solution of the crimes which they are called upon to investigate. There is a reason for this. They place more stress on their lines of information and their acquaintance with criminals and criminal methods. . . . 'What help,' they say, 'will science be in catching pick-pockets, bunco men, swindlers, and other types of criminal offenders?'"). Bunco is "[a] swindle perpetrated by means of card-sharping or some form of confidence trick," from "banca, a card-game similar to monte." 2 OXFORD ENGLISH DICTIONARY 654 (2d ed. 1989).

In its early days, forensic science was thus at some distance from the academy, but at the same time, it did not comfortably inhabit the universe of law enforcement either. To some extent, this interstitial set of relationships, in which forensic science is neither fish nor fowl, still affects both practice and culture.

Whatever the origin of current relationships, university-based academic programs in forensic science can usefully assist the creation of a research culture. We have already described the benefits of encouraging a small number of research-based forensic science programs. A tension often exists between disciplinary training—for example, in statistics, or psychology—and inherently interdisciplinary training, as forensic science education necessarily would be. But it need not be all or nothing. Some researchers in forensic science should come from disciplines like computer science, psychology, chemistry, biology, and statistics. But there is no reason why others might not come from university-based forensic science programs able to provide sophisticated training and access to disciplinary experts in the relevant subdisciplines. This dual-track approach to forensic research is likely to be more effective than either solely discipline-based research, or solely forensic-science-department-based research standing alone. We also recognize that the Forensic Science Educational Program Accreditation Commission (FEPAC) has to date accredited thirty undergraduate and graduate programs in North America. More importantly, FEPAC requires some amount of research at the graduate level for accreditation.[127] While we applaud this requirement, we also believe that carefully delineated accreditation requirements can feed a research culture but cannot necessarily create one.

Another, perhaps more innovative, approach to integrating practical aspects into university programs is the development and implementation of a clinical forensic instruction program within the university system. David

127. FEPAC states:

> Each student is required to complete an independent research project. The research project shall culminate in a thesis, or written report of publishable quality. The academic program must have written guidelines for the format of the thesis or report. In addition, the results of the work shall be presented orally in a public forum for evaluation by a committee.
>
> The research shall be conducted in an environment conducive to research and scholarly inquiry, and shall provide the opportunity for faculty and students to contribute to the knowledge base of forensic science, including research directed at improving the practice of forensic science.
>
> A committee of at least three individuals to include faculty, forensic practitioners and others with specialized knowledge will evaluate the project. At least one member of the committee must be external to the department housing the academic program.

FORENSIC SCI. EDUC. PROGRAMS ACCREDITATION COMM'N, AM. ACAD. OF FORENSIC SCIS., ACCREDITATION STANDARDS § 5.3.2.4 (2010), *available at* http://aafs.org/sites/default/files/pdf/FEPACStandards072410.pdf.

Stoney has previously drawn parallels between the current state of educational practices in modern forensic science and the state of medical practice education between 1870 and 1926.[128] Stoney showed how institutions such as Johns Hopkins Medical School significantly benefited from a new innovation: the teaching hospital. Students cared for patients and discussed cases with their clinical instructors. Students learned by trying it for themselves, engaging in actual practice, rather than just watching instructors or listening to lectures. Critical thinking and concrete problemsolving ability were valued over memorization. The teaching hospital became a center of instruction, learning, and research. The teaching hospital model is similar to the *atelier* method of art instruction,[129] and both approaches have something significant to recommend to forensic science.

"Teaching forensic laboratories" would not be difficult to imagine within a university system. These laboratories could take cases from both the prosecution and defense. They could do initial analysis or perhaps could be available to reanalyze evidence. They would benefit from a lack of institutional attachment to law enforcement or structural partiality to one adversarial side.[130] Students, working under the care and instruction of trained practitioners, could learn from real-world cases and face a realistic but challenging array of circumstances. These teaching laboratories could also provide a place for investigating and assessing emerging research and techniques before their general dissemination to state and local forensic laboratories. These teaching laboratories could also be a useful site for conducting research on validation, as well as on bias and other human factors.

Teaching laboratories could also relieve traditional forensic science laboratories of much of the burden of training. Students emerging from a clinical instruction program would accumulate significantly more useful training and experience than current novice applicants for jobs in crime laboratories. Under the current system, successful applicants undergo lengthy training programs, some as long as one to three years. Inevitably, some trainees are poorly suited for the positions, or they discover that the profession is not for them. The costs of this wash-out are high, since several years and tens of thousands of dollars

---

128. *See* David A. Stoney, *A Medical Model for Criminalistics Education*, 33 J. FORENSIC SCI. 1086 (1988).

129. This method of art instruction takes its name from the French word for "artist's studio." An artist trains a small number of students in the skills and techniques associated with creating some form of representational art, starting with more basic forms and progressing through more complex methods. *See* RICHARD LACK, ON THE TRAINING OF PAINTERS: WITH NOTES ON THE ATELIER PROGRAM 67–71 (1969).

130. Indeed, perhaps a procedure could be developed in which a party could, in certain circumstances, request that a court require such an impartial laboratory to analyze disputed evidence.

have often been invested in the process before the mismatch between trainee and profession becomes clear.[131] A clinical forensic instruction laboratory would reduce this inefficient hiring and training practice. Moreover, trainees who had gained experience through a training laboratory model might gain broader exposure to a richer variety of real-world circumstances and possibly even enhanced critical thinking abilities from their hands-on experiences.

C. Improving the Culture of Forensic Science Journals

To improve the research culture of the pattern identification sciences, some changes to the current approach to journals and publications are sorely needed. First, all forensic science journals should insist upon a full-fledged commitment to research norms. Publication in any journal that is not indexed by at least some of the major indexing services should, in a sense, not even count as publication. Peer review should be serious, blind, and carried out by individuals well qualified to assess the research merits of any given article. While non-research-oriented practitioners can play a valuable role in peer review as well, evaluations by those with the necessary qualification to assess the merits and execution of any given study should dominate the criteria for acceptance. It should also go without saying that concerns about whether a given set of findings comports with practitioners' (or researchers') expectations and desires should not affect publishing decisions. To be sure, some findings are more interesting or surprising than others, and this may legitimately affect evaluations of a given article. But the fact that a research result might alienate or irritate practitioners ought not to affect publication decisions.

The pattern identification disciplines would also benefit from a genuine flagship journal that crosses between forensic science itself and broader research paradigms. Perhaps the *JFS*, which is already a legitimate and respected research vehicle, can play this role. However, the pattern disciplines make up a small part of *JFS* publications, and the *JFS* does not especially focus on the intersections of forensic science with other disciplines. Whether a new flagship journal focusing on pattern evidence would be feasible is a difficult question. In an ideal world, such a journal would link in equal quantities to other forensic sciences and also to other academic fields, like statistics, the physical sciences,

131. In a case study involving one laboratory, poorly designed hiring procedures led to attrition costs estimated at roughly $850,000, and estimated lost productivity of nearly $5 million (because of the loss of sixteen employees). These cost estimates did not include the costs associated with recruiting, selection, or training. *See* W. Mark Dale & Wendy B. Becker, *A Case Study of Forensic Scientist Turnover*, FORENSIC SCI. COMM. (July 2004), http://www2.fbi.gov/hq/lab/fsc/backissu/july2004/research/2004_03_research04.htm.

cognitive psychology, and computer science. How to create such a flagship journal is not obvious, but one place to start would be with a high-powered and interdisciplinary editorial board that reaches broadly into ancillary disciplines as well as including leading members of the forensic science research community.

D. Using Scientific Standards to Guide Casework

Another set of suggestions focuses on efforts to use conventional scientific standards to guide casework. One key example is "sequential unmasking."[132] Analysts should have access to all the domain-relevant information they need to conduct their inquiry, but they should be shielded from domain-irrelevant matters unless or until those matters affect the analysis. A fingerprint examiner, for example, likely does need to know the surface from which a print was lifted. A fingerprint examiner does not need to know, however, about the suspect's confession or his three prior convictions for similar crimes. A document examiner cannot escape seeing the content of the document being analyzed; however, she need not be told broader aspects of the prosecution's theory of the case.

Sequential unmasking creates protocols that protect examiners from these kinds of biasing information.[133] From a research culture perspective, sequential unmasking offers two significant benefits. First, it protects examiners from materials and knowledge that might otherwise have a biasing effect on their evaluation.[134] The enormous literature on bias and cognition suggests the value of providing such a shield.[135]

Sequential unmasking has another benefit as well. Because it requires practitioners to think carefully about what information is domain-relevant and what is not, and why, sequential unmasking also encourages precisely the kind of careful attention to the relationship between evidence and warrant that a research culture demands. The very process of thinking hard and justifying the

132. *See* Dan E. Krane et al., *Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation*, 53 J. FORENSIC SCI. 1006 (2008). The basic idea of sequential unmasking (without the use of that label) was set out in Risinger et al., *supra* note 49, at 50–51.

133. Krane et al., *supra* note 132. To be sure, sequential unmasking may also increase costs by requiring an additional layer of personnel to assess what information is domain-relevant and to ensure that non-domain-relevant information is stripped from the materials the examiner receives. But given the strong evidence in other fields of the biasing effects of context information, the onus arguably ought to be on the forensic practitioner community to show why these costs are not worth incurring.

134. Dror & Rosenthal, *supra* note 32, at 902–03; Dror & Charlton, *supra* note 32, at 612; Risinger et al., *supra* note 49, at 45. Note that domain-relevant information may also generate bias, for example, when a DNA examiner looks at a mixture already knowing the suspect's profile.

135. For an overview of some of this literature and its highlights, see generally Risinger et al., *supra* note 49.

inclusion or exclusion of certain kinds of information from an examiner's purview will be a meaningful step toward the instantiation of a research culture.

Implementing blind proficiency tests in the stream of casework would be another way to make casework better comport with scientific principles for the production of knowledge. Controlled, double-blind studies are the gold standard in medicine.[136] In a double-blind study, practitioners and patients do know that they are participating in a research study, but their potential interpretive biases and expectation effects are reduced because they do not know if they are receiving the medication being tested or a placebo. Similarly, proficiency test subjects (and those administering the test) ideally ought not to know when they are pursuing ordinary casework and when they are undergoing a proficiency test.

Improving documentation practices in order to increase transparency is another step to incorporate scientific standards. While the particular degree of documentation may appropriately vary with the complexity of the comparison, documentation should be both thorough and transparent. If a fingerprint examiner, for example, finds additional minutiae on a latent print after beginning the comparison process, this back-and-forth reasoning should be clearly noted. Similarly, an examiner should indicate the degree of confidence (for example, high, medium, low) in the existence of minutiae or striations or handwriting features in the disputed exemplar in advance of undertaking any comparison. While careful documentation is no substitute for the empirical research needed to establish the power and the limits of various techniques, it can usefully clarify an examiner's reasoning process and the basis for any conclusions, and may offer some protections from the potential biasing effect of the comparison process.[137]

E. Enhancing the "Science" in the Scientific Working Groups (SWGs)

Guidelines and standards for forensic practice in a great many forensic disciplines are developed and recommended by entities known as Scientific Working Groups (SWGs), funded by the Department of Justice. Most of these working groups, which have emerged over the past twenty years, operate under

---

136. *See*, *e.g.*, John Concato, et al., *Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs*, 342 NEW. ENG. J. MED. 1887 (2000); Henry Sacks et al., *Randomized Versus Historical Controls for Clinical Trials*, 72 AM. J. MED. 233 (1982).

137. Glenn Langenburg & Christophe Champod, The GYRO System—A Recommended Approach to More Transparent Documentation (July 9, 2010) (draft), *available at* http://projects.nfstc.org/ipes/presentations/Langenburg_GYRO-System.pdf. *See generally* Interpretation Chapter, *in* NIST/NIJ WORKING GROUP REPORT ON HUMAN FACTORS IN FRICTION RIDGE IDENTIFICATION (forthcoming 2011).

the auspices of the FBI laboratory. They were designed to develop best practices, create appropriate technical standards, and improve communications both within and among various forensic disciplines.[138] Scientific Working Groups exist for firearms and toolmarks (SWGGUN), friction ridge analysis study and technology (SWGFAST), imaging technology (SWGIT), DNA (SWGDAM), shoeprint and tire tread evidence (SWGTREAD), drug analysis (SWGDRUG), as well as for a number of other forensic disciplines. These organizations have provided important venues for consensus building, policy development, and knowledge dissemination.

However, despite the scientific label in the name of the working groups, SWGs have a rather tenuous relationship with research science. Indeed, some of them previously went by other names. For example, SWGDAM, the FBI's DNA advisory group, used to be known as TWGDAM, the *technical* working group on DNA analysis and methods.

What should a scientific working group worthy of the name look like? Certainly a legitimate scientific working group would necessarily include practitioners who could inform the group about best and current practices in the discipline as well as practical constraints that operate within that area. These participants would be critical to the proper grounding and anchoring of a forensic science working group. However, these nonresearcher practitioners should make up only a minority of the group's total members. The major focus of SWGs should be to ensure that all recommendations for methods and practices are grounded in research and validated. When insufficient research exists, SWGs should determine what research is most critical to assess standards or best practices. Given these purposes, the bulk of the membership in scientific working groups should be scientists who have a relevant research background. Indeed, some members should be scientists outside the forensic discipline of the SWG and some should come from outside of forensic science entirely. These members will offer fresh perspectives and help avoid the danger of excessive buy-in to current practices simply because they are both known and familiar. The workings of the SWG would thus be driven by scientists and scientific considerations along with thoughtful input from the practitioners who would contribute to the formulation and help to operationalize the SWG's recommendations. If SWGs were organized in this fashion, they would help create and perpetuate a research

138. For basic information about the SWGs, see *Scientific Working Groups*, FED. BUREAU OF INVESTIGATION, http://www.fbi.gov/about-us/lab/swgs (last visited Jan. 28, 2011). Several SWGs operate from locations other than the FBI. SWGRUG operates out of the DEA, and SWGSTAIN operates out of the Midwest Forensic Resource Center.

culture and ensure that forensic science fields use recommended methods and processes based on scientific principles and informed by scientific research.

While the membership of the current SWGs varies, most of them are substantially more practitioner-led than what we have just described. The working group on friction ridge analysis study and technology (SWGFAST), for example, is one of the stronger SWGs. Several current members have serious and significant research interests. But they form only a small minority of the total membership. On the one hand, given the paucity of research opportunities and the structure of forensic science, this limited proportion of research-oriented members is only to be expected. But in a research culture one would expect—and insist—that the standard-setting, guideline-creating, policy body for any given field be structured so as to ensure that its decisions are based upon data and research, not simply the result of a two-thirds vote from a practitioner-dominated working group.[139] To be sure, practitioner-led SWGs may often reach appropriate, thoughtful, and perhaps even research-based conclusions, but they also risk being guided by and influenced by populist practitioner pressures. To be worthy of their name, SWGs need to make certain that scientific findings and an appreciation for a research culture drive decisions.

The National Institute of Standards and Technology (NIST) may take over the SWGs in the near future; we hope that such a move will incorporate a major restructuring of how the SWGs operate. If such a move is merely a change in their funding source from the DOJ to NIST without significant structural changes, SWGs may be useful as sounding boards for leading practitioners, but they will continue to have little to do with a research culture.

F. Access to Data

Another needed dimension for a robust research culture is access to data and test subjects. Participation in the research enterprise must obviously be balanced against a laboratory's other needs, and a laboratory may be unable to participate in every research project asked of it. However, access to data—exemplars and databases—should not be limited to practitioners at a given laboratory. With appropriate precautions for protecting confidentiality and the necessary input of Institutional Review Boards, forensic laboratories, as well as institutions like the FBI and state and federal criminal justice authorities,

139. For SWGFAST's bylaws, as an example, see *Bylaws*, SCIENTIFIC WORKING GROUP ON FRICTION RIDGE ANALYSIS, STUDY & TECH. (Sept. 15, 2009), http://www.swgfast.org/Resources/Bylaws_3.2-Corrected.pdf.

should make data available to qualified researchers to the maximum extent possible.[140]

To create incentives for providing this access, participating as research subjects ought to become an accreditation requirement for forensic labs. Just as many law schools have implemented pro bono requirements for students, the American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLD-LAB) should require that every laboratory devote a given number of hours to participation in research. The details of how to structure such a requirement could be worked out in a variety of ways. Perhaps every employee should be allowed a modest number of paid work hours for participation as a "test subject" in the research study of her choice; or perhaps laboratories should create more structured systems for participation. Whatever the details, the point is to create workable mechanisms to encourage research participation by sometimes wary laboratories. To be clear, laboratories themselves would not necessarily spearhead these research projects. Rather, analysts would be made available as test subjects, consulting on the feasibility of certain research endeavors, providing feedback on what research questions would have practical payoff for laboratories, and creating partnerships with researchers both from within and from outside the forensic sciences.

G. Managing the Tension Between an Adversarial Culture and a Research Culture

The fact that the pattern identification fields and other forensic sciences are embedded within the legal system has made it difficult for a research culture to flourish. Numerous commentators (and the NAS Report) have criticized the institutional connections between the police, the prosecutors, and the crime laboratories. Indeed, the NAS Report, like some scholarship that preceded it, explicitly calls for making crime laboratories independent of these other domains.[141]

Clearly structural risks of both bias and partisanship stem from the institutional location of crime laboratories. Several scandals have illustrated the

---

140. This has been an ongoing issue in DNA analysis as well. For example, researchers have unsuccessfully endeavored to access anonymized DNA profiles from the United States National DNA Index system, controlled by the FBI. *See* D. E. Krane et al., *Time for DNA Disclosure*, 326 SCIENCE 1631 (2009). Yet, some researchers have had access to databases for other countries. *See* David H. Kaye, *Trawling DNA Databases for Partial Matches: What Is the FBI Afraid of?*, 19 CORNELL J.L. & PUB. POL'Y 145, 161–65 (2009).

141. NAS REPORT, *supra* note 8, at 183–84; Paul C. Giannelli, *The Abuse of Scientific Evidence in Criminal Cases: The Need for Independent Crime Laboratories*, 4 VA. J. SOC. POL'Y & L. 439, 470–73 (1997); Roger Koppl, *How to Improve Forensic Science*, 20 EUR. J.L. & ECON. 255, 258 (2005).

dangers raised by forensic scientists who may feel pressured to provide prosecutors with what they are seeking.[142] Partisanship is a serious and long-recognized danger for all kinds of expert witnesses,[143] and operating as part of the institutional apparatus of law enforcement may make practitioners unconsciously partisan.[144] Additionally, the strong institutional links to police investigators may compromise efforts to protect examiners from access to unnecessary, and potentially biasing, contextual information about the case. To be sure, there may also be benefits from the current institutional location, ranging from possible efficiency gains from police authority over forensic science, to motivational gains for forensic scientists who may benefit psychologically from being part of law enforcement. And, of course, any institutional location has its own set of costs and benefits that would need to be compared to the current set.[145]

Most, but not all, of us believe that institutional separation of laboratories from the law enforcement apparatus would be tremendously beneficial for reducing the dangers of partisanship and fostering a research culture. However, most, but not all, of us also believe that even if this is indeed a worthy and highly desirable goal, it is also unlikely to be realized in the near future. One small but constructive step toward creating at least a modicum of psychological distance between laboratories and the implicit (or, sometimes explicit) pressures from law enforcement would be a requirement that all laboratories perform a certain quantity of defense-side work, enabling analysts to gain experience in a different role vis-à-vis the adversary system.[146]

However, the problematic dynamics of adverarialism and their potentially pathological effects on a research culture go beyond the sometimes-too-cozy prosecutor-police-forensic-science relationship. The dynamics of the courtroom and of the adversarial process itself can create significant incentives for analysts to resist the collection of information or the production of data that might

142. *See, e.g.*, Locke & Neff, *supra* note 5.

143. *See, e.g.*, Samuel R. Gross, *Expert Evidence*, 1991 WIS. L. REV. 1113; Jennifer L. Mnookin, *Expert Evidence, Partisanship, and Epistemic Competence*, 73 BROOK. L. REV. 1009 (2008).

144. David E. Bernstein, *Expert Witnesses, Adversarial Bias, and the (Partial) Failure of the* Daubert *Revolution*, 93 IOWA L. REV. 451, 456 (2008).

145. *See* Dror, *supra* note 25, at 101–02.

146. At present, not only do most state laboratories not regularly conduct testing for the defense, but policies about whether state laboratory workers can consult for the defense in other jurisdictions vary. One recent controversy in Minnesota illustrates the depth of adversarial norms. When a medical examiner consulted for the defense in a case in another county, the prosecutor in her home county complained to her boss, causing the medical examiner to fear for her job. While the prosecutor later apologized and was reprimanded for his behavior, the incident captures the conceptual partisanship frequently seen in the field—the notion that state forensic science workers are tied to the prosecution. *See* Joy Powell, *Dakota County Prosecutor Reprimanded by State Board*, MINNEAPOLIS–ST. PAUL STAR TRIBUNE, May 19, 2009.

assist their adversary or weaken their own credibility. If any documented error is likely to haunt an examiner on every subsequent cross-examination, there may be little motivation to identify or audit mistakes. If difficult proficiency tests would potentially provide extensive fodder for defense attorneys, why would examiners risk shooting themselves (or the prosecutors with whom they work) in the foot by attempting to determine the limits to their own abilities?

We do not have any simple fixes for this set of structural difficulties, but we offer two suggestions. First, we would suggest that laboratories consider extending something akin to *Brady* duties to examiners themselves. Under *Brady v. Maryland*,[147] prosecutors have an ethical duty to report exculpatory evidence to defense attorneys. *Brady* has already been extended to information in the possession of agents of the prosecution such as the police,[148] and there is no reason that this should not apply to forensic scientists. While a forensic scientist may have a legal duty to disclose exculpatory evidence to a prosecutor, courts have not held *Brady* duties to extend directly from the forensic scientist or police to the defense. What would be the consequences of an ethical obligation of forensic scientists to disclose directly to the defense any exculpatory findings or any inter-laboratory disagreement regarding conclusion or interpretation? Perhaps more robust reporting requirements, in which an analyst routinely discloses any interpretive disagreement within her laboratory report, would be a simpler means to achieve a similar goal. The purpose of either a disclosure requirement or enhanced reporting norms is in part to increase the degree of perceived and subjectively felt independence from law enforcement, even if no formal institutional realignment takes place. At a minimum, a research culture should mean clear and robust expectations about transparency and documentation: Reports should carefully detail steps taken, findings reached, and internal disagreement (if any) about the results or the interpretation.

A second idea worth considering is whether there ought to be a protective evidentiary privilege that attaches to self-critical investigation and analysis in at least some circumstances. This presents an extremely difficult question of balancing competing goals. Creating a privilege that protects a laboratory from having to disclose what it learns through the investigation of an error may lead to much better error investigation that may in turn reduce future errors. But in the particular case, this benefit would come at the expense of keeping highly relevant, potentially exculpatory material from defendants. Although some courts have recognized a self-critical analysis privilege in the medical peer review context (which faces structurally similar issues, though

147. 373 U.S. 83, 87 (1963).
148. *See, e.g.*, Kyles v. Whitley, 514 U.S. 419 (1995).

typically in a civil rather than criminal setting), it is quite unlikely that courts would extend it to the criminal domain, in significant part because of the criminal defendant's constitutional due process right to exculpatory information.

Nonetheless, it may be worth considering whether there are any feasible mechanisms through which defendants' legitimate (and, in some instances, constitutionally mandated) need for information could appropriately be balanced against efforts to promote self-analysis and research. It is difficult to imagine a privilege that would protect a laboratory from the disclosure of an error in actual casework. But what if a laboratory wanted to test its examiners' proficiency on difficult and close nonmatches? Should it be able to protect itself from having to report their results? Should a researcher be protected from having to identify the laboratories that participated in a study? Or which individuals achieved what results? To what extent should some kind of research privilege protect both researchers and laboratories in order to remove one major impediment to cooperation, when the results do not directly implicate any particular case or defendant?

Finally, we believe the fear that admitting imperfections might significantly harm jurors' understanding and appreciation of the pattern identification sciences may be largely chimerical. It is not clear that jurors would substantially discount conclusions from forensic science examiners even if they were presented with information quantifying error rates greater than zero, even if they knew that this particular examiner had made an occasional mistake on proficiency tests, and even if they knew that a so-called match did not necessarily mean that every other human being (or bullet, or tool) in the world could be excluded as a potential source. Certainly mitochondrial DNA evidence—which cannot ever, standing alone, individualize, because maternal relatives share the same mitochondrial DNA—can significantly contribute to a successful prosecution. Particularly in those cases in which the pattern identification evidence was combined with other probative evidence suggesting guilt, it is hardly obvious that these caveats with regard to the pattern identification evidence would have any significant impact on juror reasoning.[149] And in those rare cases

149. This is, of course, an empirical question. We can, however, make some guesses from the literature on how jurors weigh expert testimony. *See, e.g.*, Valerie P. Hans et al., *Science in the Jury Box: Jurors' Comprehension of Mitochondrial DNA Evidence*, LAW & HUM. BEHAV. (forthcoming 2011); David H. Kaye et al., *Statistics in the Jury Box: How Jurors Respond to Mitochondrial DNA Probabilities*, 4 J. EMPIRICAL LEGAL STUD. 797 (2007); Jonathan J. Koehler et al., *The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?*, 35 JURIMETRICS 201 (1995); Dawn McQuiston-Surrett & Michael J. Saks, *Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact*, 59 HASTINGS L.J. 1159 (2008); Dawn McQuiston-Surrett & Michael J. Saks, *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear*, 33 LAW & HUM. BEHAV. 436 (2009); Dale A. Nance & Scott B. Morris, *An Empirical Assessment of Presentation Formats for Trace*

where the pattern identification evidence largely stands alone, perhaps a greater degree of skepticism would be epistemologically warranted.

CONCLUSION

Our purpose in writing this Article has been to bring together a group of practitioners and academics who have all spent time thinking hard about forensic science, to see if we could find consensus about how to improve the field. Although many of us inhabit overlapping intellectual and professional circles, we did not all know each other beforehand, and we come from a variety of different intellectual traditions and locations. This project therefore began as something of an experiment. Its origin was at a conference held at UCLA in February 2010, on the one-year anniversary of the release of the NAS Report. After the public symposium, sponsored by the UCLA School of Law's Program on Understanding Law, Science, and Evidence (PULSE), this group of coauthors gathered for an intense, day-long brainstorming session.

As we discussed, outlined, and argued, we discovered that our views had more in common than one might have expected.[150] Indeed, we found that in many important respects, our views of what forensic science most needed significantly converged.

We all believe that the NAS Report got far more right than it got wrong. We all believe that many forms of forensic science today stand on an insufficiently developed empirical research foundation. We all believe that forensic science does not yet have a well-developed research culture. These disciplines, in our view, need to increase their commitment to empirical evidence as the basis for their claims. Sound research, rather than experience and training, must become the central method by which assertions are justified. While there can indeed be a legitimate role for experience-based claims of knowledge, such claims need to be both put forward with appropriate epistemic modesty and assessed through feedback mechanisms. The answer to the question "How well can you do what you say you can do?" is more properly answered by blind

---

Evidence With a Relatively Large and Quantifiable Random Match Probability, 42 JURIMETRICS J. 403 (2002); Dale A. Nance & Scott B. Morris, *Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence With a Relatively Small Random-Match Probability*, 34 J. LEGAL STUD. 395 (2005); Jason Schklar & Shari Seidman Diamond, *Juror Reactions to DNA Evidence: Errors and Expectancies*, 23 LAW & HUM. BEHAV. 159 (1999); Jonathan J. Koehler, *If the Shoe Fits They Might Acquit: The Perceived Value of Shoeprint Testimony* (Sept. 13, 2010) (unpublished manuscript, on file with UCLA Law Review).

150. Jerry Kang also participated in our day-long session and used MindManager to "map" our conversation in real time. Both our brainstorming process and the drafting of this Article were greatly assisted by his tremendous mindmapping skill.

proficiency tests than by reference to experience or training. The forensic sciences need to increase their commitment to transparency along a variety of dimensions—from increasing the documentation provided in complex cases, to more readily sharing data with researchers, to increasing access to protocols and standard procedures, to acknowledging and learning from errors. In addition, the pattern and impression fields, as well as other forms of forensic science, need to develop and sustain an ongoing critical and reflective stance, in which yesterday's truths can be revisited tomorrow.

We have offered a number of suggestions for ways to develop and improve a research culture in these fields, but we are frankly more confident in our diagnosis than in our specific suggestions for possible cures. We are, however, unanimous in hoping and believing that this is a rather special historical moment, a time when cultural change in forensic science—even perhaps, a genuine "paradigm shift"[151]—is possible. Perhaps, just perhaps, the very fact of our writing this Article together provides a small piece of evidence that this change has already begun.

151. Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892, 895 (2005) ("[W]e envision a paradigm shift in the traditional forensic identification sciences in which untested assumptions and semi-informed guesswork are replaced by a sound scientific foundation and justifiable protocols."). For the classic discussion of paradigm shifts in science, see THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (2d ed. 1970).

The Problem of Partisan Experts and the Potential for Reform Through Concurrent Evidence

By David Sonenshein and Charles Fitzpatrick*

I. INTRODUCTION

In the United States, expert witnesses are selected, paid, and prepared by the parties to the litigation. In this Article, we will explain why this system often leads to biased and partisan testimony from experts and explore several possible options for reform.

In Part II, we will examine how the American adversarial system and its allowance for party payment and preparation of expert witnesses lead to an inevitable and unavoidable danger of partisan bias in expert testimony. We will also explain why this problem is exacerbated by a lack of jury competence in evaluating expert testimony and has led to contempt for experts amongst lawyers, judges, and those professional groups from which experts are often chosen. Finally, we will show that the lack of uniform ethical guidelines for testifying experts opens the door to biased testimony.

In Part III, we will demonstrate that, although the problem of expert bias has long been recognized, past efforts for reform have been either ignored or proven inadequate. We will focus a large portion of this section on the widespread call for an increase in the use of court-appointed experts as well as some of the reasons that this practice has not become more prevalent.

In Part IV, we will explore the possibility of American courts adopting a system akin to those in the civil law universe of continental Europe, including the regimes used in Germany, France, and Italy, whereby expert witnesses are appointed by the courts. These systems aim for neutral and independent expert testimony. We will argue that the use of court-appointed neutral experts—with some modifications which acknowledge the imperatives of the adversarial system—might provide a model for effective reform. We will then turn to the adversarial systems of England, Canada, and Australia. Despite sharing a legal tradition with the United States, we will show that the English system offers relatively little that could improve the use of expert witnesses in this country, but that the procedures used in the Canadian provincial system offer some intriguing ideas for reform. Finally, in Part V we will suggest that the American legal system should also consider adoption of recent reforms in Australia, which continue to allow experts to be chosen by the parties but focus on cooperation between competing experts.

In Part VI, we conclude by arguing that the Canadian provincial and Australian national models offer a proven mechanism for diminishing expert bias within the adversarial tradition in the United States while also avoiding the difficulties which have prevented effective reform in the past.

II. THE PROBLEM OF EXPERT BIAS

The role of the expert witness in the American system is explained by Federal Rule of Evidence 702 which states, in part, "[i]f scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education may testify thereto in the form of an opinion or otherwise."[1] As the system has become inundated with more complex litigation and cases of a more technical nature, the use of expert witnesses has increased.[2] In fact, the use of experts has come to be recognized as a necessity for parties to put on the strongest possible case and is actually explicitly required in some instances.[3]

---

* David A. Sonenshein is the Jack E. Feinberg Professor of Law at Temple University Beasley School of Law. Charles Fitzpatrick is an associate with the law firm Blank Rome LLP. The views expressed in this Article are those of the authors and not necessarily those of Blank Rome LLP or of any of its clients.

1. FED. R. EVID. 702.
2. M. Neil Browne et al., *The Perspectival Nature of Expert Testimony in the United States, England, Korea, and France*, 18 CONN. J. INT'L L. 55, 66 (2002).
3. *Id.* (citing Stephen D. Easton, *"Yer Outta Here!" A Framework for Analyzing the Potential Exclusion of Expert Testimony Under the Federal Rules of Evidence*, 32 U. RICH. L. REV. 1, 8 (1998) (explaining that according to the law of some states, a plaintiff bringing a medical or other professional malpractice action must present admissible expert testimony about the standard of care and the

Nonetheless, many commentators who have examined the use of expert witnesses in the United States have concluded that "[e]xpert witnesses in court are often not deserving of our confidence. Their conclusions cannot be relied upon, and their words cannot be trusted."[4] They argue that an adversarial trial system which allows the use of experts who are selected, paid, and prepared by the parties to the litigation has a built-in danger of producing biased and partisan testimony.[5] As described by one such commentator, the American use of expert witnesses is akin to "hiring the cookie monster to guard Girl Scout cookies; the temptation to take a little bite here and another there is too great."[6]

There is a danger that a party-selected expert will unconsciously lose some degree of objectivity and slant his testimony in favor of that party which hired him as he prepares for and becomes enmeshed in the case.[7] This problem was famously described by John H. Langbein in the *University of Chicago Law Review*:

> At the American trial bar, those of us who serve as expert witnesses are known as 'saxophones.' This is a revealing term, as slang often is. The idea is that the lawyer plays the tune, manipulating the expert as though the expert were a musical instrument on which the lawyer sounds the desired notes. I sometimes serve as an expert . . . and I have experienced the subtle pressure to join the team—to shade one's views, to conceal doubt, to overstate nuance, to

---

defendant's failure to meet this standard)). *See also* Mitsubishi Elec. Corp. v. Ampex Corp., 190 F.3d 1300, 1313 (Fed. Cir. 1999) (emphasizing the importance of experts by stating that "it is well recognized that the persuasiveness of the presentation of complex technology-based issues to lay persons depends heavily on the relative skills of the experts").

4. Jennifer L. Mnookin, *Expert Evidence, Partisanship, and Epistemic Competence*, 73 BROOK. L. REV. 1009, 1010 (2008).

5. *See id.* at 1010–11 (explaining the perceived dangers of parties using experts). *See also* Samuel R. Gross, *Expert Evidence*, 1991 WIS. L. REV. 1113, 1125–53 (1991) (discussing the difference between the selection and preparation of lay witnesses versus expert witnesses and how the American system of adversarial fact finding is flawed with respect to expert witnesses).

6. Browne, *supra* note 2, at 65 (quoting Michael D. Bayles, *Professional Power and Self-Regulation*, 5 BUS. & PROF. ETHICS J. 26, 35–36 (1986)).

7. Mnookin, *supra* note 4, at 1010–11.

> downplay weak aspects of the case that one has been hired to bolster. Nobody likes to disappoint a patron; and beyond this psychological pressure is the financial inducement. Money changes hands upon the rendering of expertise, but the expert can run his meter only so long as his patron litigator likes the tune. Opposing counsel undertakes a similar exercise, hiring and schooling another expert to parrot the contrary position. The result is our familiar battle of opposing experts. The more measured and impartial an expert is, the less likely he is to be used by either side.[8]

Worse, some experts may be willing to sell their opinions and credentials to anyone who will meet their price.[9] For example, in the 1970s, a psychiatrist nicknamed "Dr. Death" was called to testify for the prosecution in more than fifty sentencing hearings.[10] According to published accounts, Dr. Death always testified that the defendant would be violent in the future regardless of whether he had actually examined the defendant.[11] Although his conclusions were contrary to the position of the American Psychiatric Association, which concluded that no psychiatrist can accurately predict the potential of a defendant for future acts of violence, he was repeatedly and consistently asked to testify by the prosecution.[12]

The problem of skepticism toward experts is not a new one. An 1870 study reported that judges and juries were attaching less and less significance to scientific testimony based on the "surprising facility with which scientific gentlemen will swear to the most opposite opinions upon matters falling within their domain."[13] In a more recent example, an expert physician, selected and paid by the

8. John H. Langbein, *The German Advantage in Civil Procedure*, 52 U. CHI. L. REV. 823, 835 (1985).

9. Mnookin, *supra* note 4, at 1011.

10. L. Timothy Perrin, *Expert Witness Testimony: Back to the Future*, 29 U. RICH. L. REV. 1389, 1442 (1995) (citing Joseph R. Tybor, *Dallas' Doctor of Doom*, NAT'L L.J., Nov. 24, 1980, at 1).

11. *Id.*

12. *Id.*

13. Ted Golan, *Revisiting the History of Scientific Expert Testimony*, 73 BROOK. L. REV. 879, 916 (2006) (quoting I.T. Hoague, *Expert Testimony*, 5 AM. L. REV. 227, 228 (1871)).

plaintiff in a case regarding silicone breast implants, claimed his testimony was based on personal examination of more than 4,700 women.[14] It was later discovered that the expert had not actually conducted the examinations himself, but rather had paid a medical student to conduct the examinations on his behalf.[15]

The potential for biased testimony is compounded by the inability of lay jurors to accurately assess the validity of technical evidence presented by an expert.[16] The result has been "a systematic distrust and devaluation of expertise"[17] amongst attorneys, judges, and even the experts themselves.[18]

A. Payment of Experts Leads to Biased Testimony

In order to protect the truth-finding purpose of a trial, it is generally illegal to pay for a witness's testimony.[19] Lay witnesses "may not be paid anything beyond nominal witness fees and expenses. All common law jurisdictions, however, allow experts to contract for special fees for their services and testimony."[20] Expert

14. Sven Timmerbeil, *The Role of Expert Witnesses in German and U.S. Civil Litigation*, 9 ANN. SURV. INT'L & COMP. L. 163, 171 (2003) (citing George C. Harris, *Testimony for Sale—The Law and Ethics of Snitches and Experts*, 28 PEPP. L. REV. 1, 2–3 (2000)).

15. *Id.*

16. Mnookin, *supra* note 4, at 1014.

17. Langbein, *supra* note 8, at 836.

18. *See* Gross, *supra* note 5, at 1135 (asserting that the present system of obtaining expert witnesses breeds contempt in judges, lawyers, and more qualified experts).

19. *See* 18 U.S.C. § 201(c)–(d) (2009) (prohibiting a party from giving or receiving anything of value "for or because of the testimony under oath or affirmation given or to be given by such person as a witness upon trial" but permitting "the payment or receipt of witness fees provided by law, or the payment, by the party upon whose behalf a witness is called and receipt by a witness, of the reasonable cost of travel and subsistence incurred and the reasonable value of time lost in attendance at any such trial, hearing, or proceeding, or, in the case of expert witnesses, a reasonable fee for time spent in the preparation of such opinion, and in appearing and testifying").

20. Gross, *supra* note 5, at 1129. Payment of expert witnesses is justified because it is unrealistic to imagine that experts would devote their time and energy preparing for and testifying at trial pro bono on a regular basis. Mnookin, *supra* note 4, at 1011.

witnesses in major cases may be able to charge the parties $500 to $1,000 an hour.[21]

Because experts are paid, there is now a thriving industry of individuals who make their living testifying for or consulting with litigants.[22] Thousands of individuals, spanning a vast array of subjects, advertise their availability to testify on behalf of litigating parties as expert witnesses.[23] Many advertise their services in legal publications.[24] Others seek employment through services which act as referral firms and headhunting agencies for experts.[25] These firms put attorneys in touch with experts for a fee and advertise their own ability to find an expert that fits the needs of the lawyer.[26]

In most cases, any minimally qualified practitioner of the expert discipline at issue is eligible to testify.[27] Thus, the parties are free to select their experts from a large pool of potential witnesses. There is no limit on the number of potential expert witnesses with whom a party can consult before trial. As described by one commentator, "I was told in one case, where a person wanted a certain thing done, that they went to sixty-eight people before they

21. Adam Liptak, *Experts Hired to Shed Light Can Leave U.S. Courts in Dark*, N.Y. TIMES, Aug. 12, 2008, at A1. *See* STEFANO MAFFEI, COMPARATIVE STUDY ON EXPERT WITNESS IN COURT PROCEEDINGS, 1, 62 (2010) (providing an example of a contract between an attorney and an expert for initial case review and consultation).

22. Gross, *supra* note 5, at 1131; Lewis A. Kaplan, *Experts in the Courthouse: Problems and Opportunities—Remarks at the Milton Handler Antitrust Review*, 2006 COLUM. BUS. L. REV. 247, 248 (2006); Perrin, *supra* note 10, at 1411–12.

23. Perrin, *supra* note 10, at 1411.

24. *See* Gross, *supra* note 5, at 1131 (referring to an issue of the *National Law Journal* that includes a page of advertisements for expert witness testimony).

25. Perrin, *supra* note 10, at 1411–12 (citing Ann Kates Smith, *Opinions with a Price*, U.S. NEWS & WORLD REP., July 20, 1992, at 64 (claiming that the Technical Advisory Service for Attorneys (TASA) listed more than 18,500 experts on its roles)).

26. *See* Gross, *supra* note 5, at 1132 (describing, as an example, the Medical Quality Foundation, which promises that its 1,150 medical experts are board certified, eminently qualified, and can effectively double the monetary value of a case).

27. *Id.* at 1127 (citing as an example United States v. Viglia, 549 F.2d 335 (5th Cir. 1977) (allowing a physician with no particular expertise in the area in question to testify on the use of a drug alleged to treat obesity)).

found one . . . . [T]herefore, I have always the greatest possible distrust of scientific evidence of this kind."[28]

This so-called "expert shopping" gives the party the opportunity to hire the expert based almost exclusively on the content and manner of their testimony.[29] Lawyers will often give little consideration to whether the individual is the most knowledgeable or most respected in the field.[30] Rather, some argue that lawyers "shop for experts, ultimately choosing the one that talks right, looks right, has the right credentials, and will work with the lawyer in the development of her opinions."[31]

Attorneys prize those experts who have the best testimonial manner and appealing credentials, but avoid those who look bad, speak poorly, or have insufficiently impressive diplomas.[32] Experts who come across as measured and impartial are unlikely to be chosen by either side.[33] Thus, "[a] fool with a small flair for acting and mathematics might be a more successful witness than say, Einstein."[34]

Unfortunately, the correlation between the qualities attorneys seek in expert witnesses and the truth is eliminated when they are sold like commodities.[35] Because an individual's success in being chosen as an expert is not dependent on her knowledge of the issue, but rather on her ability to testify persuasively to the viewpoint that the paying party wants to hear,[36] there are substantial incentives for the professional expert to advocate positions that are not supported

---

28. Mnookin, *supra* note 4, at 1010.

29. Gross, *supra* note 5, at 1133, 1143.

30. Perrin, *supra* note 10, at 1415.

31. *Id.* *See also* Sanja Kutnjak Ivkovic & Valeri P. Hans, *Jurors' Evaluations of Expert Testimony: Judging the Messenger and the Message*, 28 LAW & SOC. INQUIRY 441, 470 (2003) (emphasizing that clarity of the expert's presentation is critically important in terms of influencing the jury).

32. Gross, *supra* note 5, at 1133.

33. Langbein, *supra* note 8, at 835.

34. Perrin, *supra* note 10, at 1415–16 (quoting GERRY SPENCE, WITH JUSTICE FOR NONE 270 (1989)). *See also* Gross, *supra* note 5, at 1133 (quoting a litigator as stating, "Usually, I like my expert to be around 50 years old, have some gray in his hair, wear a tweedy jacket and smoke a pipe . . . . You must recognize that jurors have prejudices and you must try to anticipate these prejudices . . . . Some people may be geniuses, but because they lack training in speech and theater, they have great difficulty conveying their message to the jury").

35. Gross, *supra* note 5, at 1134.

36. Mnookin, *supra* note 4, at 1012.

by available research and data.[37] According to one prominent trial lawyer,

> [Expert witnesses] supply information that can salvage a lost cause or turn a winning case into a loser by purposely misleading the jurors. A lawyer who presents false evidence can be held in contempt of court. Yet there is nothing wrong with using professional opinion that puts the jury in a trance and leads them off on a tangent. . . . Expert witnesses sell their services like anyone else in the legal profession, and the best in the field can sound convincing defending either side of an argument. Their function is to snow the jury.[38]

The process of being repeatedly retained by the parties with which they are identified can further solidify expert biases.[39] "Obviously, [an expert witness] is highly motivated out of self interest to develop relationships with lawyers because those relationships are the expert's lifeblood. The more effective the expert is in advancing the lawyer's case, the greater the likelihood the expert will be retained [and paid] again."[40]

---

37. Perrin, *supra* note 10, at 1413.

38. *Id.* at 1441 (quoting ROY GRUTMAN & BILL THOMAS, LAWYERS AND THIEVES 128 (1990)).

39. Gross, *supra* note 5, at 1132–33 ("It is common . . . in many jurisdictions, for some physicians to be identified as 'plaintiffs' doctors' and others as 'defendants' doctors.' Once these labels become known, these doctors are retained repeatedly by the sides with which they are identified, a process which solidifies their biases.").

40. Perrin, *supra* note 10, at 1413. *See also* Trower v. Jones, 520 N.E.2d 297, 300 (Ill. 1988) (upholding cross examination of expert witness about how much the expert made annually for services related to rendering expert testimony). In *Trower*, the Illinois Supreme Court described the financial incentives for an expert witness as follows:

> [W]e reach our decision based on an appreciation of the fact that the financial advantage which accrues to an expert witness in a particular case can extend beyond the remuneration he receives for testifying in that case. A favorable verdict may well help him establish a "track record" which, to a professional witness, can be all-important in determining not only the frequency with which he is asked to testify but also the price which he can demand for such testimony.

There is a general consensus that the "expert witness industry has grown exponentially and with it the misuse of experts as mere partisan mouthpieces."[41] As one commentator notes, "The marketplace for experts cannot, therefore, be trusted to produce reliable information."[42]

In addition to the decreased emphasis on pure scientific truth, the payment of experts leads to economic costs. First, commercially available testimony often leads to a "cancelling effect."[43] This results when each side pays an expert and the dueling opinions negate each other.[44] Since each side spent money on an expert, with little to no effect on the jury's ability to discern the truth, there is a net economic loss.[45] Second, paid experts can lead to costly incorrect verdicts.[46] This results either when the technically incorrect side prevails or when a compromise results in an incorrect amount of damages.[47] Third, there is a potential two-pronged "distrust externality."[48] The first distrust externality is that even the most careful and trustworthy experts may be ignored in a system inundated with biased experts.[49] The second externality affects experts and includes their opportunity costs and any discomfort they have becoming part of the legal process.[50] Finally, the fourth cost is the "misallocation of expertise."[51] Because the experts do not fully internalize the costs of becoming an expert witness, they do not choose to become experts at an economically optimal level.[52] This

---

Id.

41. Perrin, *supra* note 10, at 1469. *See also* Gross, *supra* note 5, at 1132 (explaining that when experts become "repeat performers" whose inclinations are known and who are hired because of the testimony they give, "professional partisanship" becomes a problem); Mnookin, *supra* note 4, at 1015 (recognizing the risk of partisanship that is posed by expert testimony as an "old problem").

42. Mnookin, *supra* note 4, at 1012.

43. Jeffrey L. Harrison, *Reconceptualizing the Expert Witness: Social Costs, Current Controls, and Proposed Responses*, 18 YALE J. ON REG. 253, 263 (2001).

44. *Id.*

45. *Id.*

46. *Id.* at 264.

47. *Id.*

48. *Id.* at 265.

49. *Id.* at 265–66.

50. *Id.* at 266.

51. *Id.*

52. *Id.*

may artificially inflate the incentive to become an expert witness and draw many academics away from scholarly work and toward expert testimony.[53]

B. Partisan Preparation of Experts Leads to Biased Testimony

Partisan preparation of witnesses is a traditional feature of the adversarial system of fact finding employed in the United States.[54] It is the responsibility of lawyers that operate within an adversarial system to present only the evidence that favors his side, and to do so intelligently.[55] In fact, the rules of evidence assume that a lawyer will be able to anticipate the answer to each question he asks on direct examination.[56] Thus, before a trial, an attorney will generally meet with each witness he intends to call to the stand, discuss the witness's testimony, explain the process of examination, go over lists of questions and answers, anticipate lines of likely cross-examination, and offer advice in the form of testimony.[57]

The process of working with the attorney seems designed to bias the witness and produce partisan testimony.[58] Many experts admit that they face significant pressure from the party that hires them to skew their testimony.[59] In one published survey of experts,

---

53. *Id.* at 266–67.

54. *See* Gross, *supra* note 5, at 1136 ("Partisan preparation is inherent in our method of adversarial fact finding.").

55. *Id.*

56. *Id.* at 1137 (citing as examples FED. R. EVID. 611(c) and 104(a)(2) (stating that a direct examiner may not ask leading questions, but must be able to make an offer of proof advising the court of the anticipated answers to his or her questions)).

57. *Id.* at 1136.

58. *Id.*

59. *See* Joseph Sanders, *Expert Witness Ethics*, 76 FORDHAM L. REV. 1539, 1577–78 (2007) (stating that expert witnesses are often pushed to manipulate their testimony and skew their statements towards the position of the side that hired them). Sanders quotes one expert who described his first preparation sessions with a trial attorney as follows:

> [The attorney] asked me a question about whether the belt was on or not, the lap belt. And I said, "Well, could have been. But then, it may not have been." Woo, rockets went off. "What do you mean? You're my expert in this case, and you say it 'could be' or 'couldn't be?' Look, I'm going to tell you. The other side doesn't waffle. They pick one view. And

seventy-seven percent agreed with the statement, "[l]awyers manipulate their experts to weaken unfavorable testimony and strengthen favorable testimony."[60] Fifty-five percent of experts interviewed in the same survey agreed that "[l]awyers urge their experts to be less tentative."[61] Additionally, many lawyers focus expert witness preparation on potential jury biases instead of the most relevant factual evidence in the case.[62]

This potential bias from preparation is compounded by the fact that the adversarial nature of the legal system tends to draw experts who are "marginal" in that they are more willing than most of their colleagues to give opinions based on less than overwhelming evidence.[63] Also, the legal system's reliance upon verbal formulas to encapsulate key concepts may "rigidify" ideas that scientists tend to treat much more flexibly.[64] In sum, through the process of expert-shopping and pre-trial preparation, many experts have become partisan players whose testimony prioritizes confidence and clarity over scientific integrity.

Economic reward incentivizes experts to fulfill this role. As previously discussed, if the expert values her role as a witness, she will be motivated to work with the attorney who has called on her, as careful preparation and close collaboration are likely to increase the satisfaction of that attorney with the testimony and in turn make it more likely that the expert will be retained as a witness in future

---

> they will push that view. And they will make their case in front of a jury. And there will be no misunderstanding. There will be no gray area. They will take a position one way or the other and make it stick. Now, they don't have any other course of action. That's their life. They make their living going in front of juries and making statements, whether they have facts to back them up or not. Now you, you can go back to designing cars. You have another career. They don't. You better start thinking like they do."

Id. at 1578.

60. *Id.* at 1577 (citing Daniel W. Shuman et al., *An Empirical Examination of the Use of Expert Witnesses in the Courts—Part II: A Three City Study*, 34 JURIMETRICS J. 193, 201 (1994)).

61. *Id.*

62. Jody Weisberg Menon, *Adversarial Medical and Scientific Testimony and Lay Jurors: A Proposal for Medical Malpractice Reform*, 21 AM. J.L. & MED. 281, 286 (1995).

63. Susan Haack, *Irreconcilable Differences? The Troubled Marriage of Science and Law*, 72 LAW & CONTEMP. PROBS. 1, 16 (2009).

64. *Id.* at 19.

cases.[65] Further, because an expert is paid for this time, an expert witness is generally all too happy to spend significant time working to fully develop her testimony.[66]

The nature of in-court expert testimony also requires that far more time be spent in preparation than is the case for lay witnesses.[67] Because an expert, unlike a lay witness, is called to testify about matters that are beyond the common knowledge of the jury, and because the lawyer must ask questions in the way most effective for his client's case, there is a special need for collaboration as well as close and careful preparation between the expert and the attorney.[68]

Extensive preparation of an expert witness is also necessary because of the cross-examination and rebuttal she is likely to face. In addition to facing all the modes of impeachment available in the cross-examination of a lay witness,[69] experts are likely to be questioned about their training, their observations, their opinions, and the bases for their opinions.[70] These lines of questioning invite "abusive cross-examination. Since each expert is party-selected and party-paid, he is vulnerable to attack on credibility regardless of the merits of his testimony."[71] During cross-examination, an expert is likely to face accusations that his testimony has been bought rather than based on his studied professional opinion.[72]

In addition, experts hired by the opposing side are likely to criticize their opponent's methods and conclusions.[73] Typically, "an expert witness risks being attacked as not merely wrong, but

---

65. Gross, *supra* note 5, at 1138.

66. *Id.*

67. *See* Perrin, *supra* note 10, at 1417 (arguing that experts need more preparation than lay witnesses to testify because they lack personal knowledge of the case at issue).

68. Gross, *supra* note 5, at 1138.

69. FED. R. EVID. 611.

70. Gross, *supra* note 5, at 1139.

71. Langbein, *supra* note 8, at 836.

72. *Id.* ("A mode of attack ripe with potential is to pursue a line of questions which, by their form and the jury's studied observation of the witness in response, will tend to cast the expert as a 'professional witness.' By proceeding in this way, the cross-examiner will reap the benefit of a community attitude, certain to be present among several of the jurors, that bias can be purchased, almost like a commodity.").

73. Gross, *supra* note 5, at 1139. *But see* Section II.C, *infra*, describing how the jury does not have access to much of the information that would reveal the source of expert witness bias.

unqualified, ignorant, incompetent, biased, misleading or silly. Moreover, the attack is not directed to some passing observation but to her profession, her life's work."[74]

Expert witness preparation "pushes the expert to identify with the lawyers on [their] side of the lawsuit and to become a partisan member of the litigation team."[75] In addition to developing a sense of camaraderie with the attorney with whom they are working during the preparation process, experts are dependent on that lawyer to ensure their success and to protect them from attacks from the opposing side.[76] "Lawyers use their power of preparation to shape the expert's opinions. The lawyer decides what information the expert receives, what issues the expert testifies about, and, in some instances, the words the expert uses in stating her opinions."[77]

C. *Lack of Jury Competence and Limited Access to Expert's Background Contribute to the Problem of Expert Bias*

Often, juries are left to choose between the testimonies of experts from the opposing sides which have each offered conflicting opinions. Because jurors lack the scientific knowledge needed to evaluate the expert's testimony, "it can be very difficult for a jury to determine which expert opinion is correct."[78] After conflicting testimony was given in a recent case in a New York state court, the judge told the *New York Times* that "[t]he two experts were biased in favor of the parties who employed them . . . and they had given predictable testimony. 'The two sides have canceled each other out.'"[79]

In this situation, the advantage is likely to go to the party whose expert the jury found most appealing.[80] David L. Bazelon, a

---

74. Gross, *supra* note 5, at 1139.

75. *Id.*

76. *Id.*

77. Perrin, *supra* note 10, at 1418 (citing Shuman, *supra* note 60, at 202 (stating that 65% of responding lawyers believed that experts are willing to be coached about how testimony should be presented)).

78. Timmerbeil, *supra* note 14, at 169.

79. Liptak, *supra* note 21, at A1.

80. Langbein, *supra* note 8, at 836 ("If the experts do not cancel each other out, the advantage is likely to be with the expert whose forensic skills are the more enticing.").

former chief judge for the United States Court of Appeals for the District of Columbia,[81] has noted that because judges and juries lack the independent expertise to form an opinion on technical or scientific issues, a litigant's "success may depend on the plausibility or self-confidence of the expert, rather than his professional competence."[82] As noted by many trial lawyers, "[j]uries often find it hard to evaluate the expert testimony on complex scientific matters . . . and they tend to make decisions based on the expert's demeanor, credentials, and ability to present difficult information without condescension. An appealingly folksy expert . . . can have an outsize effect in a jury trial."[83] In sum, lack of knowledge and experience in the area of testimony may make a jury incapable of rendering a wise judgment on the subject.[84]

The qualities in a witness which fact-finders find most convincing such as verbal fluency, ease of manner, the appearance of humility, and stellar credentials, have little relation to the truth.[85] Further, they are an inaccurate indicator of partisan bias.[86] Thus, it is possible for an appealing yet untruthful and biased expert to sway the jury in favor of the party for which he has testified.[87] As stated by one noted scholar on the subject, "[t]he concept that the . . . jury can detect a fraud is absurd."[88] This is not a desirable result for a system which strives to discover the truth.

Additionally, many facts about the expert and the circumstances of the expert's preparation for the case are hidden from the jury. Though the risk of expert bias is well known, "some of the bias-fostering elements of the system have been allowed to

---

81. *Biographical Directory of Federal Judges*, FEDERAL JUDICIAL CENTER, http://www.fjc.gov/public/home.nsf/hisj (last visited Sept. 9, 2012).

82. Timmerbeil, *supra* note 14, at 169 (quoting John Basten, *The Court Expert in Civil Trials—A Comparative Appraisal*, 40 MODERN L. REV. 174, 174 (1977)).

83. Liptak, *supra* note 21, at A1.

84. Menon, *supra* note 62, at 283.

85. Gross, *supra* note 5, at 1134.

86. Mnookin, *supra* note 4, at 1014 ("Without epistemic competence, the jury has no choice but to rely on proxies as secondary indicia of bias, and these may often be either inaccurate or difficult to evaluate.").

87. *See id.* (arguing that because the jury lacks the expertise needed to evaluate expert testimony, it is unlikely to detect partisan bias).

88. Perrin, *supra* note 10, at 1469 (quoting Michael H. Graham, *Discovery of Experts Under Rule 26(b)(4) of the FRCP: Part Two, an Empirical Study and a Proposal*, 1977 U. ILL. L.F. 169, 189 n.44).

remain hidden and therefore prosper and grow."[89] Traditionally, the lone tool for exposing these biases has been cross-examination.[90] However, in order to fully reveal an expert's bias, an opposing attorney must have access to the materials the expert consulted in forming an opinion.[91] In some cases, the opposing attorney also must have access to communications between the expert and the attorney who prepared him or her to testify.[92] Without this information, some scholars argue, the jury can never be adequately informed about how the expert arrived at the opinion he or she presents in court. For instance, the jury should know when an overly zealous attorney shaped an expert's testimony.[93]

One commentator's solution would amend the federal rules to require full disclosure of attorney-expert communication as well as copies of all the information relevant to the development of expert opinion.[94] The resulting system would give jurors a clearer picture of any potential bias held by the expert including any influence the attorney had over them during trial preparation.[95] The commentator also argues that the common fears against full disclosure, the disruption of the work product doctrine and the attorney-client privilege, are unfounded.[96]

D. *Perceived Bias Breeds Contempt for Expert Witnesses*

The potential for bias in the testimony presented by expert witnesses in the American judicial system is widely recognized in

---

89. Stephen D. Easton, *Ammunition for the Shoot-Out with the Hired Gun's Hired Gun: A Proposal for Full Expert Witness Disclosure*, 32 ARIZ. ST. L.J. 465, 472–73 (2000).

90. *Id.* at 473.

91. *Id.*

92. *Id.* at 473–74.

93. *Id.* at 519.

94. *Id.* at 527–28. *See id.* at 544–49 (detailing proposed rule 26 amendments). *See also* Stephen D. Easton, *That Is Not All There Is: Enhancing Daubert Exclusion by Applying Ordinary Witness Principles to Experts*, 84 NEB. L. REV. 674, 714–15 (2006) (arguing that once an attorney shares his opinions with an expert, those opinions are likely to affect an expert's testimony and should therefore be disclosed).

95. Easton, *supra* note 89, at 549.

96. *See id.* at 576–608 (arguing that work product doctrine would not be disrupted because the attorney should only share his thoughts that he would not mind the jury knowing as well).

the legal community.[97] Malvin Belli, the self-proclaimed "King of Torts"[98] once said, "If I got myself an impartial witness, I'd think I was wasting my money."[99] This potential for bias has led to disdain for experts amongst lawyers and judges alike.[100]

The disdain for expert testimony extends beyond the legal professions. Some of the most respected members of other fields believe that colleagues who agree to testify face strong pressure to become partisans for the side that calls them, are treated in a demeaning manner while providing their evidence at trial, and that the evidence they present is poorly used.[101] As a result, some well-qualified experts refuse to be witnesses, "leaving the field to those with fewer scruples or fewer options."[102]

Some professional groups openly acknowledge the conflicts faced by testifying experts. For example, the ethical guidelines of the American Academy of Psychiatry and Law note,

> The adversarial nature of most legal processes presents special hazards for the practice of forensic psychiatry. Being retained by one side in a civil or criminal matter exposes psychiatrists to the potential for unintended bias and the danger of distortion of their opinion. It is the responsibility of psychiatrists to minimize such hazards by acting in an honest manner and striving to reach an objective opinion.[103]

---

97. *See* CAROL KRAFKA ET AL., FED. JUDICIAL CTR., JUDGE AND ATTORNEY EXPERIENCES, PRACTICES, AND CONCERNS REGARDING EXPERT TESTIMONY IN FEDERAL CIVIL TRIALS 21 (2002) (noting that judges and lawyers agreed in separate surveys in 1998 and 1999 that the biggest problem with expert testimony in civil cases was that "[e]xperts abandon objectivity and become advocates for the side that hired them").

98. Browne, *supra* note 2, at 70.

99. *Id.*

100. Gross, *supra* note 5, at 1135 ("One of the most unfortunate consequences of our system of obtaining expert witnesses is that it breeds contempt all around. The contempt of lawyers and judges for experts is famous.").

101. *Id.*

102. *Id.* at 1135–36.

103. *American Academy of Psychiatry and the Law Ethics Guidelines for the Practice of Forensic Psychiatry, IV, Honesty and Striving for Objectivity*, Commentary (May 2005) *available at* http://www.aapl.org/ethics.htm (explaining that "[p]sychiatrists should not distort their opinion in the service of the retaining party. Honesty, objectivity and the adequacy of the clinical evaluation may be

Nonetheless, "in many professions, service as an expert witness is not generally considered honest work . . . ."

In addition, experts themselves admit that they sometimes find their work compromising.[104] A psychologist who has testified as an expert on behalf of the defendant in a lawsuit recently told the New York Times: "After you come out of court you feel like you need a shower. They're asking you to be certain of things you can't be certain of."[105]

E. *Lack of Uniform Ethical Guidelines Creates Environment for Biased Testimony*

Often, individuals who agree to serve as an expert witness do not have a definitive ethics resource to consult. The ABA's Model Rules of Professional Conduct and Model Code of Professional Responsibility merely forbid attorneys from paying experts a contingent fee and note that experts have a duty to the court above the client.[106] The Model Code also stresses that expert witnesses should "testify truthfully and should be free from any financial inducements that might tempt them to do otherwise."[107] Though these guidelines offer some restrictions on expert testimony, it is important to remember that it is the attorney who is subject to the

called into question when an expert opinion is offered without a personal examination"; noting that "[c]ontingency fees undermine honesty and efforts to attain objectivity and should not be accepted. Retainer fees, however, do not create the same problems in regard to honesty and efforts to attain objectivity and, therefore, may be accepted").

104. Liptak, *supra* note 21, at A1.

105. *Id.*

106. John P. Murphy, *Expert Witnesses at Trial: Where Are the Ethics?*, 14 GEO. J. LEGAL ETHICS 217, 229–30 (2000).

107. *See id.* at 229–30 (quoting MODEL CODE OF PROF'L RESPONSIBILITY EC 7-28 (1999)). *See also* MODEL CODE OF PROF'L RESPONSIBILITY EC 7-28 n. 47 (1986) ("The prevalence of perjury is a serious menace to the administration of justice, to prevent which no means have as yet been satisfactorily devised. But there certainly can be no greater incentive to perjury than to allow a party to make payments to its opponent's witnesses under any guise or on any excuse, and at least attorneys who are officers of the court to aid it in the administration of justice, must keep themselves clear of any connection which in the slightest degree tends to induce witnesses to testify in favor of their clients.")

ABA regulations and who has the ultimate burden of compliance.[108] Also, when it comes to scientific testimony, these rules offer minimal restrictions on attorneys.[109] Unless an expert explicitly states that his or her testimony will be "absolutely erroneous," a violation of these rules is extremely rare.[110]

Alternatively, some experts may be able to look to professional organizations in their field for guidance on giving expert testimony. For instance, both the American Medical Association ("AMA") and the American Psychological Association ("APA") have articulated standards for expert testimony.[111] Still, many specialty organizations do not have disciplinary measures in place for addressing unethical testimony.[112]

Even if these guidelines can help the testifying expert, they still do little to solve the ethical problems of expert testimony. Most notably, there exists a conflict between the goals of attorneys and the goals of experts.[113] Attorneys operate within an adversarial environment in which their primary goal is to convince the fact-finder of their point of view.[114] In contrast, science requires a focus

---

108. Murphy, *supra* note 106, at 230.

109. David S. Caudill, *Legal Ethics and Scientific Testimony: In Defense of Manufacturing Uncertainty, Deconstructing Expertise and Other Trial Strategies*, 52 VILL. L. REV. 953, 957 (2007).

110. *Id.* ("Whether by alluding 'to any matter that will not be supported by admissible evidence' (Rule 3.4(e)), by presenting evidence known to be false (Rule 3.3(a)(3)) or by making frivolous claims or contentions (Rule 3.1), an attorney can violate the ethical rules, but the prerequisites are, respectively, no reasonable belief in admissibility, personal knowledge of falsity and no good faith basis at all for a claim or contention. With respect to scientific testimony, unless an attorney's expert states that his or her testimony will be absolutely erroneous, it is difficult to see how these rules can be violated in practice.").

111. Murphy, *supra* note 106, at 231–34. The AMA has five factors for medical testimony: 1) "the physician is a professional with special training," 2) the physician cannot become partial, 3) the physician "should testify truthfully" and be prepared, 4) the physician must make the attorney aware of both positive and negative evidence, and 5) "the physician must not accept a contingency fee." *Id.* at 231. The APA's guidelines call for psychologists to use their professional judgment, recognize and warn about the limits of their assessments, and not use outdated procedures. *Id.* at 233.

112. Edward K. Cheng, *Same Old, Same Old: Scientific Evidence Past and Present*, 104 MICH. L. REV. 1387, 1399 (2006) (noting that only seven out of thirty-six specialty organizations surveyed had disciplinary procedures in place).

113. Murphy, *supra* note 106, at 234–35.

114. *Id.* at 235.

on evidence without the influence of striving for a particular conclusion.[115] This inherent tension will continue to hinder the use of unbiased expert testimony until the system is reformed.

III. FAILED EFFORTS AT REFORM IN THE UNITED STATES

It has long been recognized that the American tradition of an adversarial trial system which allows the use of experts who are selected, prepared, and paid by the parties to the litigation leads to inevitable questions of bias.[116] Nonetheless, efforts for reform have been either ignored or proven inadequate leaving what one commentator describes as "a systematic distrust and devaluation of expertise."[117] In his view, "[s]hort of forbidding the use of experts altogether, we probably could not have designed a procedure better suited to minimize the influence of expertise."[118]

A. *Frye and Daubert*

Under the traditional common law approach, although experts might provide overtly partisan testimony, it was believed that each side had an equal opportunity to bring truth to the forefront through use of their own experts and cross examination.[119] Under this view, "so long as parties had an equal opportunity to bring forward opposing experts, under the same rules and with the same judge as umpire, then whatever the jury made of the competing experts' reports was acceptable."[120]

This approach fell under criticism as early as the end of the nineteenth century.[121] Even with a level adversarial playing field,

115. *Id.*

116. *See generally* Jennifer L. Mnookin, *Idealizing Science and Demonizing Experts: An Intellectual History of Expert Evidence*, 52 VILL. L. REV. 763 (2007) (providing a detailed look at the views of expert evidence at the end of the nineteenth century).

117. Langbein, *supra* note 8, at 836.

118. *Id.*

119. Mnookin, *supra* note 4, at 1015.

120. *Id.* (citing Roscoe Pound, *The Causes of Popular Dissatisfaction with the Administration of Justice*, 14 AM. LAW. 445, 447–48 (1906)).

121. *Id.* (citing Mnookin, *supra* note 116, as providing a look at the historic anxieties surrounding the use of expert witnesses).

critics felt that juries lacked the competence to adequately evaluate the reliability of expert testimony because of the complicated nature of such testimony.[122] As noted by a scholar from that period, this practice had the effect of turning "expert witnesses[] into partisans pure and simple."[123]

In 1923, the Court of Appeals for the District of Colombia created what has come to be known as the "Frye test" or "general acceptance test."[124] The *Frye* court held that a scientific principle must have gained "general acceptance" in its field to form the basis of expert testimony.[125] Under this test, "it was not enough that one qualified expert believed the procedure was reliable. Instead, the court must determine that 'general acceptance' has been reached."[126]

This standard dominated the admission of scientific evidence for at least fifty years.[127] Critics of the Frye test claimed the rule was difficult to apply because it required the judge to determine: in what community should he look to see if the standard was accepted; who should be counted in that community; how should votes be counted; what constitutes "general acceptance"; how can "general acceptance" be proven; and, how should the "general acceptance" of conflicting techniques be resolved.[128] Others claimed that, in practice, the courts often ignored the required focus on "general acceptance."[129] Rather, they argued that courts only required a demonstration of adequate qualifications to support a witness's claim of expertise and rarely questioned such a claim in any meaningful

---

122. *Id.* at 1016.

123. *Id.* (quoting Pound, *supra* note 120, at 448).

124. *See generally* Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

125. *Id.* at 1014.

126. Browne, *supra* note 2, at 56–57. In *Frye*, the defendant in a murder trial attempted to offer expert testimony regarding the results of a systolic blood pressure lie detector test which indicated his innocence. *Frye*, 293 F. at 1013. The court held that because the test had not gained the requisite standing and recognition among psychological and physiological authorities, the trial court had been correct in refusing to allow defendant's expert to testify. *Id.* at 1014.

127. Browne, *supra* note 2, at 56 (citing Jon P. Thames, *It's Not Bad Law—It's Bad Science: Problems with Expert Testimony in Trial Proceedings*, 18 AM. J. TRIAL ADVOC. 545, 549 (1995)).

128. *Id.* at 59–60.

129. *See* Mnookin, *supra* note 4, at 1016 ("[I]n practice, most judges, most of the time, did not actually interrogate a proposed experts' bona fides in a detailed or rigorous way.").

way.[130] Thus, these critics maintained that the application of the Frye standard led to testimony from unethical and biased experts, allowing lawsuits to move forward and sizeable verdicts to be awarded despite the lack of scientific merit.[131]

Nonetheless, as late as 1983 the Supreme Court continued to express support for the traditional view of cross-examination as "beyond any doubt the greatest legal engine ever invented for the discovery of truth."[132] The Court continued to believe that "the adversary process [could] be trusted to sort out the reliable from the unreliable evidence" and made no further efforts to protect against partisan expert testimony.[133]

In 1993, in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the Court held that the Frye test had been superseded in federal law by the promulgation of Federal Rule of Evidence Rule 702.[134] In applying Rule 702, the Court held that trial judges should act as gatekeepers, and only allow scientific evidence which is truly based on "scientific knowledge."[135] Rather than requiring the jury to determine scientific fallacies through cross-examination, the Court directed judges to allow the jury to hear only that testimony which was deemed "reliable."[136]

The *Daubert* court identified four non-exclusive "general observations" which trial judges should utilize in assessing the reliability of scientific evidence before allowing the presentation of expert testimony at trial: (1) whether the expert's methodology has been tested; (2) whether the theory applied by the expert or the methodology utilized has been published and subject to peer review; (3) what the method's rate of error is when it has been applied and what standards of control direct the technique's operation; and,

130. *Id.*

131. *Id.* at 1017.

132. Bruce D. Black, *The Use (or Abuse) of Expert Witnesses in Post-*Daubert *Employment Litigation*, 17 HOFSTRA LAB. & EMP. L.J. 269, 270 (2000) (citing 5 JOHN HENRY WIGMORE, A TREATISE ON THE ANGLO-AMERICAN SYSTEM OF EVIDENCE IN TRIALS AT COMMON LAW § 1367 (3d ed. 1940)).

133. Barefoote v. Estelle, 463 U.S. 880, 901 (1983).

134. Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 589 (1993) (holding that the "general acceptance" test was absent from and incompatible with the Federal Rules of Evidence and should not be applied in federal trials).

135. *Id.* at 589–91.

136. *Id.* at 589.

(4) whether the theory or technique has been generally accepted in the relevant scientific or professional community.[137]

In *Kumho Tire Co. v. Carmichael*, the Court extended the rationale and factors in *Daubert* to non-scientific areas of expertise.[138] The text of Rule 702 itself was amended on December 1, 2000, to require judges to serve as gatekeepers who would exclude unreliable evidence of both the scientific and un-scientific variety when presented by an expert witness.[139] Writing for the majority in *Kumho*, Justice Breyer explained that the "objective of that requirement is to ensure the reliability and relevancy of expert testimony. It is to make certain that an expert . . . employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."[140] Thus, if the expert's chosen theory or methodology has no use outside of the courtroom or cannot be verified by reference to real world examples, it is much less likely to be admissible.[141] If the expert's principle and methodology are utilized in science, industry, or some other practical forum outside of litigation, they are likely acceptable for use in trial testimony.[142]

Thus, *Daubert* can be seen as an indirect response to the potential for biased expert testimony created by party hiring and payment of experts and a lack of jury competence.[143] Under this standard, an expert's qualifications will not be enough to allow his testimony to be admissible.[144] Nor will allowance for the presentation of conflicting evidence from an opposing expert be

137. *Id.* at 593–94.

138. Kumho Tire Co. v. Carmichael, 526 U.S. 137, 141 (1998) (holding that the trial judge's general "gatekeeping" obligation and the *Daubert* principles apply "not only to testimony based on 'scientific' knowledge, but also to testimony based on 'technical' and 'other specialized' knowledge").

139. FED. R. EVID. 702 advisory committee's notes on the 2000 amendments.

140. *Kumho Tire Co.*, 526 U.S. at 152.

141. Black, *supra* note 132, at 270–71. *See also In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 742 (3d Cir. 1994) (quoting U.S. v. Downing, 753 F.2d 1224, 1239 (3d Cir. 1985) (noting that in determining the reliability of expert testimony, courts should consider "non-judicial uses to which the scientific technique are put")).

142. ANTHONY J. BOCCHINO & DAVID A. SONENSHEIN, A PRACTICAL GUIDE TO FEDERAL EVIDENCE 114 (Anthony J. Bocchino & Zelda Harris eds., 1991).

143. Mnookin, *supra* note 4, at 1019.

144. *Id.*

enough to counter the potential for biased testimony.[145] Rather, it is hoped that the traditional adversarial tools of cross-examination and the presentation of conflicting evidence, combined with enhanced judicial scrutiny and the power to exclude testimony determined to be "unreliable," will reveal an expert's bias and provide the fact finder with the information needed to properly evaluate expert testimony.[146]

Unfortunately, application of the Daubert criteria has not adequately eliminated the potential for biased expert testimony.[147] A study released in 2002 based on surveys of judges and lawyers reveals that although "expert testimony has received increased judicial attention in the years since *Daubert* . . . problems with testifying experts have been largely unaffected by the passage of time. Judges and attorneys in the recent surveys reported frequent problems with partisan experts . . . These same issues dominated in pre-*Daubert* times."[148]

The shift to the Daubert criteria has also had several unintended consequences.[149] These include a new area of discovery into the peer-review process as well as the growing prevalence of "litigation-driven scholarship."[150] In terms of the peer-review

145. *Id.*

146. Perrin, *supra* note 10, at 1424. In Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1998), the plaintiffs claimed that a tire blowout, which resulted in death and injuries to the vehicle's occupants, was caused by a defect in the tire. The plaintiffs' case was based in large measure on the deposition of a tire failure analyst, who intended to testify as an expert that a defect in the tire's manufacture or design caused the blow out. His opinion was based upon a visual and tactile inspection of the tire and upon the theory that in the absence of at least two of four specific, physical symptoms indicating tire abuse, the tire failure of the sort that occurred here was caused by a defect. Applying the *Daubert* principles, the Court held that the trial court was correct in determining that the expert's opinion was unreliable. The Court found no indication in the record that other experts in the industry followed the experts particular approach nor any reference to articles or papers which could be used to validate the approach.

147. *But see* Cheng, *supra* note 112, at 1401–02 (noting that "[t]his transformative potential of *Daubert*, coupled with the modern willingness to accept the role of the managerial judge, may encourage a greater degree of inquisitorial thinking, opening the door to institutions like court-appointed experts and scientific tribunals").

148. KRAFKA, *supra* note 97, at 24.

149. William G. Childs, *The Overlapping Magisteria of Law and Science: When Litigation and Science Collide*, 85 NEB. L. REV. 643, 645–46 (2007).

150. *Id.*

process, publication-related documents have been subpoenaed and participants in the process have been deposed.[151] Some fear that this injection of legal procedure into the scientific research process may have an adverse effect by chilling independent research.[152]

Litigation-driven research is the phenomenon in which scientists conduct research in preparation for litigation and then subsequently submit it for publication in an attempt to bolster the likelihood that their testimony will be admitted.[153] While the potential for bias in such practice is obvious, it is important to avoid a universal ban of litigation-driven research for several reasons.[154] First, some research that is in the public interest will only be conducted in the context of litigation.[155] Thus, it is important not to dissuade such research. Second, just because research is conducted in the context of litigation, and therefore funded by a party with clear interest in the results, it is not necessarily bad science.[156] Instead, the judge and jury must evaluate each litigation-driven study independently and determine its reliability accordingly.[157]

B. Court-Appointed Experts and Other Failed Efforts at Reform

1. Science Courts and Expert Judges

In addition to the Frye and Daubert tests, there have been several other reform proposals aimed more specifically at attacking bias in expert witness testimony.[158] These efforts, however, have all proven either too difficult to implement or ineffective in practice.

Some have proposed the creation of a special "science court" consisting of experts capable of evaluating the different scientific arguments of parties and their experts.[159] This proposal has never

151. *Id.* at 645.
152. *Id.*
153. *Id.*
154. *Id.* at 670–71.
155. *Id.* at 670.
156. *Id.* at 670–71.
157. *Id.* at 671.
158. Timmerbeil, *supra* note 14, at 170.
159. *Id.* (citing Arthur Kantrowitz, *The Science Court Experiment: Criticism and Responses*, 33 BULLETIN OF ATOMIC SCIENTISTS 44, 44 (1977) ("The Science Court is intended to deal only with scientific questions of fact.")). *See also* James

been implemented due to the expense associated with the permanent hiring of qualified experts from many different scientific fields, who would need to be paid competitive salaries, and because of the need to provide such experts with a basic legal education.[160] Additionally, it is possible that that Jury Selection and Service Act of 1968, with its requirement that juries be selected at "random from a fair cross section of the community," could prevent such a system.[161]

Others have proposed establishing panels of "expert judges."[162] This proposal is seen as overly expensive and impossible to implement because of the difficulty of providing an expert capable of handling every scientific issue that may arise at trial.[163] Also, there is some concern that a specialist judge, with his or her own biases, might unduly influence the jury.[164]

2. Court-Appointed Experts

The use of court-appointed experts has received more support than any other suggested reform. In 1901, Judge Learned Hand concluded that the typical "jury is not a competent tribunal"[165] to understand the theories and techniques utilized by experts, and as such was unable to overcome the obstruction of truth by biased expert witnesses.[166] The seriousness of the problem led him to propose the creation of a system of neutral, court-appointed experts.[167] These experts would provide the jury "the final statement of what was true"[168] and recommend decisions on scientific issues.[169] Similarly, John Henry Wigmore, author of an early 20th

---

A. Martin, *The Proposed "Science Court,"* 75 MICH. L. REV. 1058, 1064–69 (1977) ("[T]he science court would be a body to pass on scientific aspects of public policy questions presented to it[].").

160. Timmerbeil, *supra* note 14, at 170.

161. Menon, *supra* note 62, at 294.

162. Timmerbeil, *supra* note 14, at 171.

163. *Id.*

164. Menon, *supra* note 62, at 295.

165. Billings Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40, 55 (1901).

166. *See* Timmerbeil, *supra* note 14, at 171 (presenting Judge Hand's argument that "expert juries" are necessary to protect against truth obstructing expert witnesses).

167. Hand, *supra* note 165, at 56.

168. *Id.* at 54.

169. *Id.*

century treatise on evidence, was moved in part by the partisanship of expert witnesses[170] to propose that "the State, not the party, shall be the one to pay his fee, and . . . the Court, not the party, shall be the one to summon him."[171]

In 1937, the National Conference of Commissioners on Uniform State Law adopted a Model Expert Testimony Act which provides for court-appointed experts.[172] The following year, the Act was endorsed by the American Bar Association Committee on the Improvement of the Law of Evidence.[173] Today, federal courts are free to appoint their own expert witnesses through authority granted in Federal Rule of Evidence 706.[174] Many other jurisdictions have

---

170. Gross, *supra* note 5, at 1189.

171. 1 JOHN HENRY WIGMORE, A TREATISE ON THE SYSTEM OF EVIDENCE IN TRIALS AT COMMON LAW § 563 (2d ed. 1923).

172. Gross, *supra* note 5, at 1189.

173. *Id.*

174. FED R. EVID. 706. The relevant provisions of Rule 706 provide:

> (a) Appointment. The court may on its own motion or on the motion of any party enter an order to show cause why expert witnesses should not be appointed, and may request the parties to submit nominations. The court may appoint any expert witnesses agreed upon by the parties, and may appoint expert witnesses of its own selection. An expert witness shall not be appointed by the court unless the witness consents to act. A witness so appointed shall be informed of the witness' duties by the court in writing, a copy of which shall be filed with the clerk, or at a conference in which the parties shall have opportunity to participate. A witness so appointed shall advise the parties of the witness' findings, if any; the witness' deposition may be taken by any party; and the witness may be called to testify by the court or any party. The witness shall be subject to cross-examination by each party, including a party calling the witness.
>
> (b) Compensation. Expert witnesses so appointed are entitled to reasonable compensation in whatever sum the court may allow. The compensation thus fixed is payable from funds which may be provided by law in criminal cases and civil actions and proceedings involving just compensation under the fifth amendment. In other civil actions and proceedings the compensation shall be paid by the parties in such proportion and at such time as the court directs, and thereafter charged in like manner as other costs.
>
> (c) Disclosure of appointment. In the exercise of its discretion, the court may authorize disclosure to the jury of the fact that the court appointed the expert witness.

also made room for the use of court-appointed experts through rules modeled on Rule 706.[175] Often, "[c]ourts use this power when a factual issue arises that would be better solved with the support of scientific knowledge and the partisan experts are not helpful due to their relationship to the parties."[176]

Despite the availability of this power, court-appointed experts are rarely used in U.S. civil courts.[177] The rule's drafters acknowledged that "actual appointment is a relatively infrequent occurrence."[178] They hoped "the availability of the procedure in itself [would] decrease . . . the need for resorting to it."[179] They postulated that "[t]he ever-present possibility that the judge may appoint an expert in a given case must inevitably exert a sobering effect on the expert witness of a party and upon the person utilizing his services."[180] Despite the drafters' original optimism, there remains a deep-seated suspicion of the reliability of expert witnesses amongst attorneys, judges, and experts themselves.[181] Thus, it seems clear that further reforms are necessary to remove the potential for biased testimony from expert witnesses.

> (d) Parties' experts of own selection. Nothing in this rule limits the parties in calling expert witnesses of their own selection.

Id.

175. Gross, *supra* note 5, at 1190 (citing JACK B. WEINSTEIN & MARGARET A. BERGER, 3 WEINSTEIN'S EVIDENCE § 706(04), at 706–40 (1988) (listing jurisdictions adopting versions of Federal Rule of Evidence 706)). In limited circumstances, some jurisdictions have also used experts in special proceedings. For instance, some experts have been used in compulsory unitization proceedings in the field of oil regulation to determine ownership interests. Still, the use of experts is limited and should likely be utilized more in processes such as compulsory unitization in which some parties believe there is procedural unfairness. Gideon Wiginton, *Addressing Perceptions of Procedural Unfairness in Compulsory Unitization by Appointing Neutral Experts*, 55 AM. U. L. REV. 1801, 1802–06 (2006).

176. Timmerbeil, *supra* note 14, at 167.

177. KRAFKA, *supra* note 97, at 19 (discussing an empirical study which found that 73.9% of federal judges had never appointed an expert in a civil case utilizing Rule 706).

178. FED. R EVID. 706 advisory committee notes on proposed rules.

179. *Id.*

180. *Id.*

181. *See supra* notes 4–18 and accompanying text.

a. Reasons Against Using Court-Appointed Experts

1. Lack of Necessity

Some deem court-appointed experts to be unnecessary in most circumstances. Court-appointed experts are useful only when the court feels the need for an independent assessment of a disputed issue.[182] The Federal Judicial Center asked eighty-one judges why they believed that the authority to appoint experts was used so infrequently.[183] Echoing Weinstein, the judges indicated that they view appointing experts as an extraordinary action.[184] The judges stated that the types of cases that require court-appointed experts are "both rare and unusually demanding, implying that appointed experts should be reserved for cases with extraordinary needs."[185] Patent, product liability, and antitrust violations were the most common cases requiring court-appointed experts.[186] Other judges responded that court-appointed experts were only necessary in response to a combination of unusual events, such as a complex technical issue combined with a need to protect a poorly represented party.[187]

A number of judges surveyed mentioned the need for a court-appointed expert when both parties' experts were in complete disagreement.[188] Such a concern is more in line with those of the Advisory Committee, whose main concern when proposing Rule 706 was the venality of experts called upon by parties.[189] But as one judge noted: "One needs a complete divergence in the views of the

---

182. JOE S. CECIL & THOMAS E. WILLGING, FED. JUDICIAL CTR., COURT APPOINTED EXPERTS: DEFINING THE ROLE OF EXPERTS APPOINTED UNDER FEDERAL RULE OF EVIDENCE 706 18 (1993).

183. Joe S. Cecil & Thomas E. Willging, *Accepting* Daubert*'s Invitation: Defining a Role for Court-Appointed Experts in Assessing Scientific Validity*, 43 EMORY L.J. 995, 1015 (1994).

184. *Id.*

185. *Id.* at 1016.

186. *Id.*

187. *Id.* at 1017. *See also* Stephanie Domitrovich, Mara L. Merlino & James T. Richardson, *State Trial Judge Use of Court Appointed Experts: Survey Results and Comparisons*, 50 JURIMETRICS J. 371, 383 Table 2 (2010) (showing that a higher percentage of state judges would appoint more experts for testimony on medicine, pharmacology, physics and economics than for any other areas).

188. Cecil & Willging, *supra* note 183, at 1018.

189. FED. R. EVID. 706 advisory committee notes on proposed rules.

parties' experts in a technically complex field. Often experts differ, but not in a crazy way."[190] Thus, it would seem that the parties' experts would have to exhibit blatant, extreme venality before a judge would consider appointing its own in response.

2. Respect for the Adversarial System

Respect for the adversarial system was cited as a major reason for the infrequent appointment of experts by the judges surveyed by the Federal Judicial Center.[191] Many of the judges professed a commitment to the adversarial system, the ability of lawyers to find qualified experts, and the ability of juries to sift through difficult evidence.[192] One judge stated:

> The lawyers are pretty good about shooting holes in each others' experts. It's generally a credibility question and the jury can sort it out . . . . In general, it conflicts with my sense of the judicial role, which is to trust the adversaries to present information and arguments. I do not believe the judge should normally be an inquisitor.[193]

Further, some scholars have argued that, by designating a witness as "court-appointed" and "impartial," the court has cloaked the witness with a "robe of infallibility."[194] These scholars are concerned that a designation of impartiality may elevate the opinion of the court-appointed expert above those of the parties' experts, regardless of their position.[195] As a result, there is concern that juries will be unduly deferential to court-appointed experts, which

---

190. Cecil & Willging, *supra* note 183, at 1018.
191. *Id.*
192. *Id.* at 1018–19.
193. *Id.* at 1019.
194. *See* Karen Butler Reisinger, *Court-Appointed Expert Panels: A Comparison of Two Models*, 32 IND. L. REV. 225, 236 (1998) (citing Elwood S. Levy, *Impartial Medical Testimony—Revisited*, 34 TEMP. L.Q. 416, 424 (1961)).
195. *Id.* at 236.

would usurp the jury's fact-finding authority.[196] However, there is no guarantee that a court-appointed expert will be unbiased. For instance, a court-appointed expert may be biased by the school of thought under which he or she was trained.[197] After all, "[s]cientists are human beings, which means that they too are subject to social and personal interests."[198]

Empirical data indicates that jurors do attach great weight to the testimony of a court-appointed expert. For example, a study of outcomes in a series of asbestos litigation revealed that jurors sided with the court-appointed expert in thirteen out of sixteen cases.[199] The judge assigned to the cases remarked: "A court's expert will be a persuasive witness and will have a significant effect upon a jury."[200] The Federal Judicial Center also surveyed judges who had appointed experts about the effect of a court-appointed expert's testimony.[201] Of fifty-eight responses, only two indicated that the result in a case involving the use of a court-appointed expert was inconsistent with the guidance given by the court's expert.[202]

Finally, many litigants and judges feel that court-appointed experts disrupt the adversarial process because it interferes with

---

196. Sophia Cope, *Ripe for Revision: A Critique of Federal Rule of Evidence 706 and the Use of Court-Appointed Experts*, 39 GONZ. L. REV. 163, 176 (2004).

197. *See* Reisinger, *supra* note 194, at 236 (explaining that when the court announces an expert "neutral," the jury is likely to believe it, but an expert might be partial simply in the way he was trained). Economics serves as a good example for this type of conflicting opinion. *See, e.g.*, Abbott B. Lipsky Jr., *Antitrust Economics—Making Progress, Avoiding Regression*, 12 GEO. MASON L. REV. 163, 168–70 (2003) (noting that "[m]any of the salient microeconomic issues underlying the leading antitrust cases of our day could be cited as examples of serious disagreements among contending schools of contemporary antitrust economists and among specific individuals and the modes of analysis they sponsor" and pointing out the difference between traditional economic antitrust models and "post-Chicago" theories).

198. Si-Hung Choy, *Judicial Education After* Markman v. Westview Instruments, Inc.*: The Use of Court-Appointed Experts*, 47 U.C.L.A. L. REV. 1423, 1450 (2000).

199. Reisinger, *supra* note 194, at 236 (citing Carl B. Rubin & Laura Ringenbach, *The Use of Court Experts in Asbestos Litigation*, 137 F.R.D. 35, 41 (1991)).

200. *Id.* at 237.

201. Cecil & Willging, *supra* note 182, at 52–56.

202. *Id.* at 52.

party autonomy.[203] When an expert witness is appointed by the court, the parties largely lose their ability to control the trajectory of that witness's testimony at trial.[204] If a trial lawyer cannot properly manage an expert's testimony, there is increased risk that the expert's examination will elicit an unfavorable response.[205] In essence, much of the debate over court-appointed experts is a turf war between judges and trial lawyers.[206] And because many judges have spent careers as trial lawyers, they seem to be more willing to cede this ground. As stated by a judge surveyed by the Federal Judicial Center: "I believe in the adversary system. I was a litigator for thirty years. I don't feel comfortable taking over the case (like a small claims court, without lawyers). I don't know why I would be better equipped than the lawyers to find a top-flight person."[207]

3. Impracticality

Many judges find Rule 706 appointments to be impractical given the constraints of litigation. For instance, judges have often found it difficult to identify the need for an expert in time to make the appointment without delaying the trial. In the Federal Judicial Center survey, just over a dozen judges stated that an effective appointment requires the court's awareness of the need early in the litigation.[208] Since parties rarely suggest that the court make an appointment, a judge may not recognize the need for an independent expert until the eve of the trial.[209] At that point, it is difficult to recruit an expert and make an appointment while fulfilling the procedural requirements of Rule 706. Most judges who have appointed experts report making the selection early in the litigation, usually at the close of discovery.[210] However, some judges have reported making an appointment during trial, or even after the trial ended.[211]

---

203. Reisinger, *supra* note 194, at 237.

204. Gross, *supra* note 5, at 1200.

205. *Id.* at 1200–01.

206. *Id.*

207. Cecil & Willging, *supra* note 183, at 1018–19.

208. Cecil & Willging, *supra* note 182, at 22–25.

209. *Id.*

210. *Id.*

211. *See, e.g.*, United States v. Weathers, 618 F.2d 663, 664 n.1 (10th Cir. 1980) (considering a court's post-trial appointment of an expert).

Another practical problem is the cost of court-appointed experts.[212] Rule 706(c) sets forth the means of compensating court-appointed experts. If the expert testifies in a criminal case, the Rule provides that the court determines the amount in whatever sum the court deems reasonable and that it is to be paid by the government as provided by law.[213] In civil cases, the cost of the experts is passed to the parties.[214] Many judges are reluctant to rely on the parties for payment and have stated that they restrict appointment to cases in which both parties consent.[215] One judge stated that lawyers find the appointment of experts in civil cases to be "hard to justify to their client when the client is paying for an expert already," especially when the court-appointed expert may hurt the client's case.[216]

In summary, many judges find Rule 706 appointments to be too impractical. Many judges recognize the need for a neutral expert too late to comply with the procedural requirements of the Rule. Even if the need is recognized, many other judges find it difficult to justify the cost when the parties are already providing their own experts.

4. The Availability of Technical Advisors

Another reason that judges may choose not to appoint an expert witness is the availability of technical advisors. The Supreme Court has recognized that trial courts have the inherent power to appoint persons to advise the court on technical matters that may arise during litigation.[217] The judge's power to appoint technical experts is not practically limited and the decision is subject to review

212. Stephanie Domitrovich, Mara L. Merlino & James T. Richardson, *State Trial Judge Use of Court Appointed Experts: Survey Results and Comparisons*, 50 JURIMETRICS J. 371, 382 (2010) (stating that the most frequently cited reason that state-level trial judges do not appoint experts was the cost to the court and the parties).

213. FED. R. EVID. 706(c).

214. *Id.*

215. Cecil & Willging, *supra* note 182, at 22.

216. *Id.*

217. *Ex Parte* Peterson, 253 U.S. 300, 306–07 (1920).

only for abuse of discretion.[218] The only limit on the power to appoint technical advisors is Article III, which forbids judges from abandoning the judicial function.[219] In *Reilly v. United States*, the First Circuit articulated several procedural safeguards, the violation of which may constitute an abuse of discretion.[220]

First, the *Reilly* court stated that technical advisor appointments should be rare.[221] Appointment should be limited to complex cases where outside expertise will hasten adjudication without disrupting the role of the judge.[222] Thus, a technical advisor should only be appointed where the trial court is faced with problems of unusual difficulty, sophistication, or complexity.[223] These relatively rare requirements mirror those that call for a court-appointed expert. Thus, the use of technical advisors is an alternative to appointing a full-fledged expert.

Still, the technical advisor's role differs from that of an expert witness in several important ways. The technical advisor's function is to act as a sounding board and to educate the judge on complex scientific and technical matters.[224] The technical advisor is not to brief the judge on legal issues, as that responsibility is tasked to the court itself.[225] If a judge allows a technical advisor to overstep these bounds, then the judge "effectively abdicates the Article III role and thereby violates the constitution."[226] Also, the *Reilly* court proposed that a judge inform the parties of the technical advisor's identity before making the appointment.[227] The judge should also give the parties an opportunity to object to the appointment.[228] Additionally, the judge should formulate a written "job description"

---

218. Robert L. Hess II, *Judges Cooperating with Scientists: A Proposal for More Effective Limits on the Federal Trial Judge's Inherent Power to Appoint Technical Advisors*, 54 VAND. L. REV. 547, 558 (2001).

219. *Id.*

220. Reilly v. United States, 863 F.2d 149, 156–57 (1st Cir. 1988).

221. *Id.*

222. *Id.*

223. *Id.* In *Reilly*, the task of estimating the future earning capacity of an infant negligently injured at birth using complex economic theories satisfied this requirement. *Id.* at 153.

224. *Id.* at 157–58.

225. *Id.* at 158.

226. Hess, *supra* note 218, at 571.

227. *Reilly*, 863 F.2d at 159.

228. *Id.*

which defines the technical advisor's role and duties.[229] The judge should then require the advisor to sign an affidavit affirming compliance with the job description.[230]

Importantly, technical advisors are also different from expert witnesses in that they are not in fact witnesses and do not give testimony.[231] As a result, many of the procedural rules governing expert witness testimony do not apply to technical advisors.[232] For instance, since they do not render expert opinion, parties do not have the right to depose, call, or cross-examine technical advisors.[233] The *Reilly* court also held that technical advisors need not submit a written report.[234]

Because technical advisors do not fall within the ambit of Rule 706 and may consult with judges ex parte without notification of the parties,[235] technical advisors would seem to be a more convenient way for judges to sift through complex technical issues. However, the lack of a clearly defined set of rules to govern the behavior of technical advisors opens the door to their potential misuse.[236]

For example, there are concerns about a judge or jury abdicating its fact-finding role to a technical advisor, potential bias of the advisor, or that a supposedly "neutral" expert could undermine the adversarial process.[237] Specifically, judges with limited expertise in a complex area could conceivably completely delegate their gatekeeping duty to a technical advisor.[238] This would be a clear violation of *Daubert's* requirement that the judge make an independent determination as to the reliability of the proffered

229. *Id. See* MediaCom Corp. v. Rates Techn., Inc., 4 F. Supp. 2d 17, 37 (Appendix B) (Mass. Dist. Ct. 1998) (outlining the technical advisor's job description).

230. *Reilly*, 863 F.2d at 159–60.

231. Weinstein & Berger, *supra* note 175 § 13.06[1].

232. Note, *Improving Judicial Gatekeeping: Technical Advisors and Scientific Evidence*, 110 HARV. L. REV. 941, 950 (1997).

233. *Reilly*, 863 F.2d at 156; *Improving Judicial Gatekeeping*, *supra* note 232, at 950.

234. *Reilly*, 863 F.2d at 158–59.

235. Hess, *supra* note 218, at 557–58.

236. *Improving Judicial Gatekeeping*, *supra* note 232, at 950. However, at least one court has endorsed the use of a written report as a proper way to define the limits of a technical advisor's role. *Id.* at 951.

237. *Id.* at 953.

238. *Id.*

evidence.[239] Alternatively, even if a judge properly consults a technical advisor, the advisor will undoubtedly have some kind of bias.[240] This is a particularly worrisome danger because technical advisors, as appointed "neutral" experts, will face less skepticism from the jury than a hired expert. Finally, technical advisors may interfere with the adversarial process by circumventing the parties and supplying evidence directly to the judge.[241]

Some ways to combat these potential dangers include allowing the parties to participate in the selection process, giving the technical advisor a written description of his job and the extent of his duties, and having the technical advisor submit a written report of his discussions with the judge to the parties.[242]

IV. FOREIGN MODELS FOR DEALING WITH EXPERT BIAS

Despite the long recognized problem of partisan testimony from expert witnesses, the implimentation of adequate reforms has proven elusive. Historically, those who have sought to modernize their own legal system have looked to methods used in other legal cultures and adopted their best practices.[243] Thus, despite the failure of past reform efforts in the United States, it may be possible to remove the problem of partisan expert testimony by looking at the models provided by other countries.

A. Typical Civil Law ("Inquisitorial") Systems

1. German Use of Court-Appointed Experts

German civil litigation operates under an inquisitorial system whereby the judge plays a very active role.[244] In Germany, the judge "controls the proceedings, examines the witnesses and is always the

239. *Id.*
240. *Id.*
241. *Id.*
242. *Id.* at 955.
243. Robert F. Taylor, *A Comparative Study in Expert Testimony in France and the United States: Philosophical Underpinnings, History Practice and Procedure*, 31 TEX. INT'L L.J. 181, 182 (1996).
244. Timmerbeil, *supra* note 14, at 163.

decision maker."[245] Through the Code of Civil Procedure (ZPO), German courts make a distinction between lay witnesses and court-appointed experts.[246] In fact, experts are not even called witnesses but rather are thought of as judges' aides.[247] The German system operates under the assumption that "credible expertise must be neutral expertise."[248] All experts are expected and required to remain neutral in regard to the parties involved in the litigation.[249]

A German judge may decide to seek the assistance of an expert through his own motion or upon the request of a party to the suit.[250] The judge selects the expert.[251] Usually experts are chosen from a list compiled by the judge himself.[252] The judge's discretion, however, is somewhat limited by ZPO § 404(2), which says that "[i]f experts are officially designated for certain fields of expertise, other persons should be chosen only when special circumstances require."[253] This provision gives priority to officially designated experts for specific fields.[254] These experts are chosen from lists of professionals qualified to serve as experts, which are compiled by official licensing bodies and state governments.[255] These individuals are sworn to render professional and independent expert assistance.[256] The bodies and governments provide judges with regularly updated lists of available experts from which the judge may choose.[257]

Although the court generally takes the initiative in nominating and selecting the expert, the court is required to use any

---

245. *Id.*

246. *Id.* at 173 (citing ZIVILPROZEBORDNUNG [ZPO] CODE OF CIVIL PROCEDURE § 402-414 (Ger.)). *See* MAFFEI, *supra* note 21, at 1, 69–70 (referencing statutes in Germany, France, and Italy pertaining to expert evidence).

247. Langbein, *supra* note 8, at 835 (citing KURT JESSNITZER, DER GERICHTLICHE SACHVERSTANDIGE 72–78 (7th ed. 1978)).

248. *Id.* at 837.

249. Timmerbeil, *supra* note 14, at 174.

250. Langbein, *supra* note 8, at 837 (citing ZPO § 404(I)).

251. Timmerbeil, *supra* note 14, at 173; ZPO § 404.

252. Timmerbeil, *supra* note 14, at 173.

253. Langbein, *supra* note 8, at 837.

254. Timmerbeil, *supra* note 14, at 174.

255. Langbein, *supra* note 8, at 837–38. *See* MAFFEI, *supra* note 21, 1, 36–40 (detailing the process of creating lists of experts in various European countries).

256. Langbein, *supra* note 8, at 838 (citing GEWERBEORDNUNG [GEWO] (CODE ON TRADE REGULATION) § 36 (Ger.)).

257. *Id.*

expert agreed upon by the parties.[258] If an expert is appointed without the consent of the parties, either party may seek the expert's recusal.[259] However, the circumstances warranting recusal are very limited.[260] ZPO § 406(1) only allows litigants to challenge an expert's appointment on the narrow grounds available for seeking the recusal of a judge.[261] Experts can only be recused when it appears that they are not neutral.[262] For example, an expert could be recused if she was a friend or relative of a party.[263] The court has authority to accept or deny such a request.[264]

Neutrality of the expert in regard to the parties is borne out through selection of experts by German judges. According to at least one noted scholar on the subject,

> [T]he most important factor predisposing a judge to select an expert is favorable experience with that expert in an earlier case. Experts thus build reputations with the bench. Someone who renders a careful, succinct, and well-substantiated report, and who responds effectively to the subsequent questions of the court and the parties will be remembered when another case arises in his specialty.[265]

When judges have not had previous experience with an appropriate expert, they then turn to the official lists.[266]

Party neutrality is also achieved through the instructions given to the expert. Although the court welcomes suggestions from the parties themselves,[267] it is the court that provides the facts that the expert is directed to investigate,[268] frames the question for the

---

258. *Id.* at 837 (citing ZPO § 404 (IV)).
259. Timmerbeil, *supra* note 14, at 174; ZPO § 406 (1).
260. Timmerbeil, *supra* note 14, at 174.
261. *Id.* (describing the reference in ZPO § 406(1) to ZPO §§ 42–45 provisions providing potential reasons to recuse a judge).
262. *Id.*
263. *Id.*
264. *Id.*
265. Langbein, *supra* note 8, at 838.
266. *Id.*
267. *Id.* at 839.
268. *Id.*

expert to address,[269] and regulates the extent to which the expert is permitted to contact the parties.[270]

Generally, the court orders an appointed expert to submit a written opinion.[271] Although the parties are usually permitted to file written comments to which the expert is asked to reply, the court can order the expert to appear at trial for further explanation.[272] If the expert is required to appear in court, the judge conducts the initial interrogation.[273] Attorneys for the parties themselves are then each provided an opportunity to pose further questions.[274] Party questioning of the expert is conducted in a polite and non-confrontational manner, quite unlike the cross-examination of experts seen in U.S. courts.[275]

Deferential questioning by the parties is due in part to a prohibition on leading questions in German civil litigation.[276] The non-confrontational atmosphere can also be attributed to the independent and unbiased position of the expert.[277] Because the expert was appointed by the court and received her instructions directly from the judge, "[a]ttacking the expert would be equivalent to criticizing the judge's authority to select and question the expert—and in German civil courts, the judge is always the decision-maker."[278] Therefore, strategic considerations demand that the parties treat the expert as neutral and independent.[279]

The parties in German civil litigation are free to seek and submit to the court the opinion of their own hired witnesses.[280] There are no rules in the ZPO dealing with these party-selected experts,[281] and German courts have held that the opinion of a party-selected expert witness does not have the same value as that of a court-appointed expert.[282] The opinion of a party-selected expert is

---

269. *Id.*
270. Timmerbeil, *supra* note 14, at 174 (citing ZPO § 404(a)).
271. *Id.* at 175.
272. *Id.*; Langbein, *supra* note 8, at 839.
273. Timmerbeil, *supra* note 14, at 175.
274. *Id.*
275. *Id.*
276. *Id.*
277. *Id.*
278. *Id.*
279. *Id.*
280. Langbein, *supra* note 8, at 840.
281. Timmerbeil, *supra* note 14, at 177.
282. *Id.* at 177–78.

not considered evidence, but rather only an assertion of the party.[283] A partisan expert's opinion can be used to discredit the court-appointed expert and can stand as grounds for engaging an additional court-appointed expert.[284] The court, however, is not required to appoint a new expert, and rarely does so in practice.[285]

Although a court must explain why its final decision does not follow the view expressed by a partisan expert,[286] courts generally discount such opinions because they view the opinions of those who have been hired by the parties and discussed the case with counsel to be biased and therefore unreliable.[287] Further, party-selected experts are not examined at trial[288] and their opinions cannot form the basis of the court's final decision.[289]

2. French Use of Court-Appointed Experts

French civil courts also utilize an inquisitorial structure whereby judges maintain strong control over all proceedings.[290] France has a long history of using expert witnesses, dating back to 1667.[291] In the French system, a judge, known as the Judge Delegate, plays an active role in the gathering of pre-trial evidence.[292] This evidence is later submitted to a separate three-judge panel for fact-finding and a ruling.[293] As in Germany, the French civil system of justice is purposely structured to maintain neutrality in the evidence provided by expert witnesses.

Expert neutrality is first advanced through appointment by the court rather than by the parties themselves. A Judge Delegate is

283. *Id.*

284. *Id.*

285. *Id.*

286. *Id.* at 179.

287. Langbein, *supra* note 8, at 840.

288. Timmerbeil, *supra* note 14, at 178.

289. *Id.*

290. *See* Taylor, *supra* note 243, at 210 (explaining the "active role in finding facts and questioning witnesses" of the Judge Delegate in the French civil system).

291. Browne, *supra* note 2, at 95 n.258. The rules have been constantly evolving and were most recently revised in the 1970s. *Id.* at 95.

292. Taylor, *supra* note 243, at 189.

293. *See id.* at 186, 189 (explaining the process by which the Judge Delegate gathers evidence, which is then submitted to a panel of three professionally trained judges for fact-finding and a decision).

authorized to appoint an expert to assist the three judge panel in its fact-finding role on either its own motion or a motion by one of the parties.[294] In addition, in some circumstances the Judge Delegate may be required by statute to appoint an expert.[295] Although each party may request appointment of an expert, such appointments are a matter of judicial discretion.[296] When such a request is granted, the Judge Delegate almost always chooses the expert without debate by the parties.[297] Further, the judge must clearly indicate the necessity of the expert's mission and the precise issues to be investigated.[298]

Neutrality is also achieved through the manner in which an expert is chosen by the court. As in Germany, French courts maintain regional and national lists of experts.[299] Although it is not required, courts generally utilize the lists in practice because of convenience and a sense of quality control.[300] In addition to the lists, the French legislature has placed strict limits on the choice of experts to be used for issues dealing with copyright and guardianship proceedings.[301] Finally, certain professions, such as chemists, physicians, and automobile mechanics are regulated by their own standards.[302] These professionals may not be chosen as experts unless they meet national qualifying standards for that particular profession.[303]

The Judge Delegate considers objectivity, competence, clarity of expression, and diligence when choosing an expert.[304] Parties may object to the chosen expert or attempt to have an expert

294. *Id.* at 192–93 (citing N.C.P.C. art. 143).

295. *Id.* at 193 n.101. As an example, Taylor explains that under C. civ. art. 126, persons who receive temporary possession of the personal property of a person who has been missing for ten years or more may petition to have a court-appointed expert visit a building for the purpose of verifying the building's condition. Further, he explains that under C. civ. art. 459, a provision governing guardianship and care of the infirm provides for the sale of business assets and/or real property of a minor, or transfer of similar assets to a partnership or corporation after the property is assessed by a court-appointed expert.

296. Browne, *supra* note 2, at 96 (citing N.C.P.C. art. 232).

297. Taylor, *supra* note 243, at 192.

298. Browne, *supra* note 2, at 98 (citing N.C.P.C. art. 265).

299. *Id.* at 96.

300. Taylor, *supra* note 243, at 195.

301. Browne, *supra* note 2, at 96–97.

302. Taylor, *supra* note 243, at 194–95.

303. *Id.* at 195.

304. *Id.*

removed from the national or regional lists.[305] Valid objections in regard to a particular case are limited to those based on prejudice, bias, or conflict of interest.[306] An expert will only be removed from one of the lists upon a demonstration of incapacity or professional misconduct.[307]

Neutrality is further reinforced through the manner in which the expert conducts her investigation. The parties are entitled to be fully informed regarding the expert's assignment, approaches, and activities.[308] Thus, once an expert has been appointed, the parties are afforded a right to request a hearing to provide comments and suggestions on how the expert should proceed.[309] "For example, the parties may request that the expert visit the scene or interview particular witnesses."[310] Although required to consider these suggestions when formulating the appropriate course of action and to provide a reason for each decision, the expert is under no obligation to adhere to the requests of the parties.[311]

The parties are given a right to be informed of the time and location of significant activities within the expert's investigation so that they may observe the expert in action.[312] Actual attendance by the parties, however, is not mandatory and the expert is under no obligation to postpone an activity due to the absence of a properly informed party.[313] Further, there is no right for the parties to attend activities by the expert which are of a purely technical nature.[314]

---

305. *Id.* at 196–97.

306. *Id.* at 196 n.135. *See also* Browne, *supra* note 2, at 97 n.280 (explaining that the grounds on which an expert can be removed are the same as those for which a judge can be removed: "(1) if the judge or his or her spouse has a personal interest in the case; (2) if the judge or his or her spouse is the creditor, debtor, heir or legator of the parties; (3) if the judge or his or her spouse is related to one of the parties closer than the fourth degree; (4) if there is a pending case between the judge or his or her spouse and one or more of the parties; (5) if the judge has ever sat as a judge of either of the parties; (6) if the judge has ever been responsible for the disposition of any property of the parties; (7) if there are any ties between the judge and his spouse and the parties or their spouses; and (8) if there is a relationship between the judge and any of the parties").

307. Taylor, *supra* note 243, at 197.

308. Browne, *supra* note 2, at 99.

309. *Id.*

310. Taylor, *supra* note 243, at 204.

311. *Id.*

312. *Id.*

313. *Id.*

314. *Id.* (citing N.C.P.C. art. 161).

The expert's mission is also defined by neutrality. The expert's guiding principle is an obligation to execute the required activity "with conscience, objectivity, and impartiality."[315] An expert in the French system is an officer of the court and never an agent of the parties.[316] An expert's mission is limited to findings on purely technical matters.[317] He is not permitted to explain the law or the legal effects of his findings.[318] The expert is required to precisely follow the court's directions, answer each question posed by the court completely and accurately, and is generally prohibited from expanding or narrowing the mandate provided by the Judge Delegate.[319]

Neutrality is further maintained by the manner in which an expert conveys his opinion to the court and the parties. Experts in the French civil system are required to file a report detailing their findings with the court[320] and may also be required to make an oral presentation to the parties.[321] When more than one expert has been utilized, the expert report must indicate areas of unanimity[322] and explain the reasons for varying opinions on points of disagreement.[323] The report must include: (1) a preamble, (2) operations of the expert, and (3) the expert's results.[324] The report is fully confidential and therefore is not available to the public.[325] Additionally, though the expert is court appointed, the judge is not required to accept the expert's findings.[326] Finally, the French civil system maintains expert neutrality through its system of compensating experts for their services. "[T]he expert may not accept any remuneration, direct or indirect, from the parties, including gifts or other benefits such as travel or food."[327] Although

---

315. *Id.* at 205 (quoting N.C.P.C. art. 237).
316. Browne, *supra* note 2, at 99.
317. Taylor, *supra* note 243, at 192.
318. *Id.* (citing N.C.P.C. art. 238).
319. Browne, *supra* note 2, at 99.
320. *Id.* (citing N.C.P.C. art. 282).
321. Taylor, *supra* note 243, at 207 (citing N.C.P.C. art. 282).
322. Browne, *supra* note 2, at 100 (citing N.C.P.C. art. 282).
323. Taylor, *supra* note 243, at 208.
324. Browne, *supra* note 2, at 99–100.
325. *Id.* at 100.
326. *Id.*
327. Taylor, *supra* note 243, at 206 (citing N.C.P.C. art. 248).

funding is actually provided by the parties themselves, it is the court which actually conveys the fee to the expert.[328]

3. Italian Use of Court-Appointed Experts

a. Civil Proceedings

As in the German and French legal systems, Italian civil proceedings are inquisitorial.[329] Consistent with the inquisitorial system, judicial appointment is the exclusive method by which experts may be appointed in an Italian civil case.[330] Although the parties may request a court-appointed expert, the decision to appoint an expert is in the sole discretion of the judge.[331]

A court-appointed expert, or "consulente tecninco d'ufficio" (CTU), is authorized by Article 61 of the Italian Code of Civil Procedure.[332] Much like American technical advisors, the CTU is an auxiliary of the judge, assisting the judge with complex technical matters that may arise during the case.[333] The CTU may not, however, usurp the judge's duty to independently determine the case's outcome.[334] Article 61 also provides that the CTU must be selected from the "Albo dei Periti," which is a register that divides individuals qualified to render technical assistance into various categories depending on their area of expertise.[335]

If a judge finds it necessary to appoint a CTU, the parties must be given leave to hire their own experts to review the work of the court's expert and to provide commentary on that expert's findings.[336] These experts, known as party consultants, are explicitly authorized by Article 201 of the Italian Code of Civil Procedure.[337]

328. Browne, *supra* note 2, at 100 (citing N.C.P.C. art. 284).

329. SIMONA GROSSI & MARIA CHRISTINA PAGNI, COMMENTARY ON THE ITALIAN CODE OF CIVIL PROCEDURE 3 (2010).

330. *Id.* at 17.

331. *Id.*

332. Rahol Kakkar & Paola Cascot, *Overview of Italian Regulation Expert Witnesses*, *available at* http://www.twobirds.com/English/NEWS/ARTICLES/Pages/2006/Overview_Italian_regulation_expert_witnesses.aspx (last visited Sept. 9, 2012).

333. *Id.*

334. *Id.*

335. *Id.*

336. Grossi & Pagni, *supra* note 329, at 17.

337. *Id.*

According to Article 201, party consultants are permitted to accompany parties to any proceeding involving the court's expert in order to comment on the results of any technical investigation or piece of evidence.[338]

If a CTU makes a technical determination in a case, the parties are also entitled to enter their own consultant's report into evidence without any prior evaluation by the judge as to its admissibility or relevance in the case.[339] A consultant's report is then treated like any other exhibit.[340] Once admitted, it is up to the judge to decide whether or not to take it into consideration when deciding the case.[341]

b. Criminal Proceedings

In 1988, the Italian Parliament approved a new Code of Criminal Procedure that was much more adversarial in nature than its predecessor.[342] Under the previous code, the entire fact-finding process was judicially controlled—the judge called witnesses, examined them ex officio, and introduced documents.[343] Under the new code, however, "evidence is received by the party's request" and, thus, each party presents his or her own case by gathering evidence, calling witnesses, and examining them.[344]

However, one aspect of the inquisitorial system that remained untouched by the new code is the treatment of expert-witness testimony.[345] Under both the old and new codes, expert witnesses are always appointed by the court.[346] Additionally, expert witnesses are examined ex officio in court.[347] Because court-appointed experts in Italian criminal cases are examined in court and provide testimonial evidence, they seem to have a more active role in aiding the judge's final decision than experts in civil proceedings.

---

338. *Id.*
339. *Id.* at 17–18.
340. *Id.*
341. *Id.*
342. Elisabetta Grande, *Italian Criminal Justice: Borrowing and Resistance*, 48 AM. J. COMP. L., 227, 244 (2000).
343. *Id.*
344. *Id.*
345. *Id.* at 244–45.
346. *Id.*
347. *Id.*

4. German, French, and Italian Systems Do Not Provide Useful Models for Reform of U.S. Courts

Although Germany, France, and Italy have structured their civil systems in a way that provides protections against the influence of partisan or impartial expert witnesses, these systems do not provide a model upon which adjustments can be made in the United States.

The use of court-appointed neutral expert witnesses may offer an effective check against the possibility of an overly partisan expert fabricating an opinion or dramatically overstating some aspects of his testimony.[348] Court-appointed experts may also be effective when there is a high degree of agreement within the relevant community on the specific question at issue in the case.[349] It is, however, less clear whether court-appointed experts offer meaningful improvement when there are disagreements between experts that reflect legitimate and well-grounded differences of opinion within a professional, technical, or scientific community.[350] In such a situation, the court-appointed expert may take a firm position, creating the appearance of consensus on a point that is actually still open to debate.[351] The attractiveness of simplifying the fact-finding role of the jury (or judge in a bench trial) is surely outweighed by the potential for a "neutral" court-appointed expert to mislead the fact finder into believing his opinion is the only one possible.[352]

Alternatively, the expert may lay out the scientific contours of the debate without taking a position.[353] This situation is no better, as it still leaves the jury in the position of trying to decide an issue on which it lacks the proper expertise.[354]

These general criticisms apply equally to the German, French, and Italian systems discussed above. Under the German

348. Mnookin, *supra* note 4, at 1026.
349. *Id.* at 1021.
350. *Id.* at 1026–27.
351. *Id.* at 1027.
352. *Id.*
353. *Id.*
354. *Id.*

ZPO, a good expert will mention other recognized scientific opinions in her report.[355] Nonetheless, if the expert supports one opinion over the others, she is obviously no longer neutral in respect to other existing opinions, and it is possible she would not mention these other opinions at all.[356]

It is nearly impossible for the parties in German litigation to adequately make up for this deficiency themselves because of the lack of an effective cross-examination.[357] Although the judge may appoint a second expert, he will do so only when he doubts the first expert's opinion.[358] Moreover, a judge is unlikely to express such a doubt.[359] Not only does the judge lack the necessary knowledge or training to effectively evaluate the expert's opinion, but he is also unlikely to abandon the opinion of an expert who he has personally appointed and likely worked with in the past.[360] It is equally unlikely that a judge will abandon his trust in an expert which he has personally appointed, in favor of an opinion rendered by an expert hired and paid for by a party to the suit.[361]

Under the French and Italian systems, a judge is never bound by the opinion of the court-appointed expert. Nonetheless, these systems, like their German counterpart, are flawed because the outcome of the trial seems to necessarily fall on the opinion of the chosen expert, regardless of whether equally valid opinions may have been given by others of equal qualifications. As described by one scholar,

> The appointee will conduct an investigation outside the courtroom, under no formal rules of evidence or relevance, at sessions to which the parties are convoked with full freedom to present their views and those of their experts or other representatives, orally or in writing. The result of the expertise is a report that in principle the judge need not accept, but in the absence of other evidence, it is difficult to see how it

355. Timmerbeil, *supra* note 14, at 180.
356. *Id.*
357. *Id.*
358. *Id.* at 182.
359. *Id.*
360. *Id.*
361. *Id.* at 175–76, 178 (empirical studies showed that the court follows the opinion of the court expert in 95% of the cases).

> could be rejected, provided that the judge is satisfied that the expert has done what he was commissioned to do and that no material procedural irregularities have been committed in the course of the expertise.[362]

Further, U.S. judges are already given powers to appoint expert witnesses that are similar to those held by German, French, and Italian courts, and yet this authority is seldom used. As noted earlier, Federal Rule of Evidence 706 codifies the common law right of judges to appoint expert witnesses to testify in court;[363] it has been used as a model for allowing court-appointed experts in many state courts.[364] Under this rule, the court may select an expert nominated by the parties or choose its own candidate.[365] Courts may also refer to lists of appointed experts in cases involving frequently litigated subjects.[366] The only limitation on this authority is that the witness must consent to the appointment.[367]

Like their German, French, and Italian counterparts, U.S. judges define the scope of a court-appointed expert's duties and will ensure that the expert's findings are reported to the parties.[368] Court-appointed experts may be directly questioned by the judge at trial

---

362. Richard W. Hulbert, *Comment on French Civil Procedure*, 45 AM. J. COMP. L. 747, 749 (1997).

363. Tahirih V. Lee, *Court Appointed Experts and Judicial Reluctance: A Proposal to Amend Rule 706 of the Federal Rules of Evidence*, 6 YALE L. & POL'Y REV. 480, 481 (1988).

364. *See* Gross, *supra* note 5, at 1189–90 ("Explicit provisions for court-appointed experts have been adopted . . . in over thirty states and territories . . . Rule 706 of the Federal Rules of Evidence is the model for a majority of the current provisions for the use of court-appointed experts.").

365. *See* FED. R. EVID. 706(a) ("The court may on its own motion or on the motion of any party enter an order to show cause why expert witnesses should not be appointed, and may request the parties to submit nominations. The court may appoint any expert witness agreed upon by the parties, and may appoint expert witnesses of its own selection."); Eric Illhyung Lee, *Expert Evidence in the Republic of Korea and Under the U.S. Federal Rules of Evidence: A Comparative Study*, 19 LOY. L.A. INT'L & COMP. L.REV. 585, 618 (1997).

366. Lee, *supra* note 365, at 619.

367. FED. R. EVID. 706(a) ("An expert witness shall not be appointed by the court unless the witness consents to the act.").

368. *Id.* ("A witness so appointed shall be informed of the witness' duties by the court in writing . . . A witness so appointed shall advise the parties of the witness' findings").

and cross-examined by the parties.[369] The parties are also free to retain and call their own experts to testify.[370] Court-appointed experts are entitled to "reasonable compensation" from court funds or payments ordered by the court from both litigants.[371]

Despite the existence of Rule 706 and its state counterparts, judges are extremely reluctant to utilize their authority to appoint their own experts.[372] To summarize the previous discussion, there have been many explanations for this reluctance. According to one commentator, "Many judges, if not most, have been trial lawyers, and they are suspicious that any expert is truly neutral." [373] Some suggest that judges fear juries will be unduly persuaded by the opinion of the court-appointed expert over those hired by the parties.[374] Others believe judges view court-appointed experts as interfering with the traditional neutrality and objectivity of the judge and the adversarial role of counsel in the American legal system.[375] Still others have suggested that many judges have either not been provided with the resources necessary to locate and select appropriate experts,[376] or are simply unaware that they have the authority to appoint experts.[377]

Regardless of the reasons for judicial unwillingness to utilize the power of expert appointment, this reluctance is pervasive

---

369. *Id.* ("[T]he witness may be called to testify by the court or any party. The witness shall be subject to cross-examination by each party."). *See* Taylor, *supra* note 246, at 212 ("Once a court-appointed expert is selected, he may be called at trial and questioned by the judge.").

370. FED. R. EVID. 706(d) ("Nothing in this rule limits the parties in calling expert witnesses of their own selection.").

371. FED. R. EVID. 706(b) ("Expert witnesses so appointed are entitled to reasonable compensation in whatever sum the court may allow. The compensation thus fixed is payable from funds which may be provided by law in criminal cases and civil actions and proceedings involving just compensation under the Fifth Amendment. In other civil actions and proceedings the compensation shall be paid by the parties in such proportion and at such time as the court directs, and thereafter charged in like manner as other costs.").

372. Lee, *supra* note 365, at 480.

373. Liptak, *supra* note 21, at A1.

374. Lee, *supra* note 365, at 495.

375. *Id.* at 496.

376. *See* Gross, *supra* note 5, at 1191 ("Why is this power of appointment so neglected? One possible answer is that Rule 706 and similar provisions are insufficiently structured; they do not define specific areas of expertise for appointment, and they do not help judges locate and select appropriate experts.").

377. Timmerbeil, *supra* note 14, at 168.

throughout the judiciary. According to a study released in 2002 based on surveys of federal judges, despite the enormous number of cases requiring expert testimony, 73.9% of judges had never appointed an expert utilizing Rule 706.[378] Given the reluctance of United States judges to utilize their already existing powers of expert appointment, there would be little benefit to adoption of practices modeled on the German, French, or Italian systems.

So what lessons can we in the United States draw from the civil law system, and what modifications could we make in the court-appointment process workable and more palatable to American lawyers and judges? First, it must be recognized that the "audience" for whom the expert performs in the United States is a jury, not a professional judge. For this reason, the appointed expert in the United States must testify, likely in a narrative form. The appointed expert must be available for the usual cross-examination, which, according to long-accepted U.S. legal doctrine, would expose any intellectual, political, or philosophical bias, as well as errors or weakness in methodology.

Second, the parties, as they can in Europe, must still be able to retain and examine in court their own experts. Hopefully, the partisan experts could find room for significant agreement on many issues when faced with the report of the court-appointed expert. If, after undergoing cross-examination from one or both parties, the court-appointed expert has more credibility than the other experts in the case, then the system has worked. As an alternative to U.S. judges arbitrarily appointing a Rule 706 expert, arbitration practice may provide a sound answer. As in the selection of arbitrators, the parties could be asked to nominate proposed court-appointed experts, providing respective lists where in the parties may agree on certain names. In addition, as in arbitration, the parties' experts could be expected to agree on a neutral court-appointed expert. Because the cost of the neutral court-appointed expert will be borne by the parties, some might argue that the cost of the neutral court-appointed expert will be prohibitive. First, we are not suggesting that the court appoint an expert under Rule 706 in every case. The court should retain the discretion to appoint when the value of the case can tolerate the additional cost and where the issues may be less than straightforward. Moreover, it would make sense that the court would

378. Krafka, *supra* note 97, at 19.

only consider making the appointment after receiving and reviewing the expert reports of the retained experts. This, in turn, might create an incentive for partisan experts to take less extreme positions.

B. Adversarial Legal Systems

Though the civil law system offers reforms which we believe can be adapted to the United States justice system, we recognize that American lawyers and judges may feel some resistance to adopting civil law systems' modes of dealing with experts in our adversarial system. Therefore it would be helpful to turn to the experience of other adversarial systems for guidance. These systems share commonalities with the American system and their practices are likely to be perceived as more easily transferable to United States courts.

1. English System

The English system, as a close analogue to the American system, would appear to provide a useful comparison. In general, opinion evidence is inadmissible unless it falls into one of three exceptions, including the expert opinion evidence exception.[379] This exception admits evidence to "[prove] matters of specialized knowledge."[380] Additionally, experts must have "competence," and the court must investigate the expert's qualifications to prove competence.[381] Still, England has an intermediate evidentiary position compared to the rules of other countries.[382] This means that the general admissibility of evidence is determined by the finder of law, while the weight of the evidence is left up to the finder of fact.[383]

Another important consideration in the English system is the relative availability of experts to the various parties. While some countries have adopted specific rules outlining the availability of experts to each party, England's rules, at least in theory, treat both

---

379. Browne, *supra* note 2, at 76–77.
380. *Id.* at 77.
381. *Id.* at 78.
382. *Id.* at 79.
383. *Id.* at 79–80.

sides equally.[384] However, forensic experts are typically associated with the prosecution; therefore, in practice, the defense tends to have fewer experts to choose from.[385] Thus, the defense typically has a disadvantage when it comes to the use of expert testimony.[386]

Beyond these general access concerns, there has been a recent history of misuse of scientific evidence in England.[387] In response, an effort to reform the system led to the creation of the "Royal Commission on Criminal Justice" to investigate the misuse of scientific evidence for the purpose of convicting the accused.[388] The Commission made several suggestions, including the establishment of a code of ethical and scientific standards for expert testimony and better instructions on the evidence during trials.[389] The Commission also explicitly rejected the idea of court-appointed experts.[390] Though the Commission's findings were illuminating, they only represented suggestions and not mandatory changes. Some legal scholars criticized the Commission for its lack of skepticism about scientific testimony.[391] The Commission's findings are part of a larger trend in English law; arguably, under the English system, the scientific expert has greater status than her counterpart in the United States because the law does not examine scientific evidence with the same caution.[392]

Finally, England's system of free evaluation of evidence causes some additional problems.[393] Because the tribunal of fact is permitted to attribute any amount of worth to the evidence, it may be misled and end up granting certain evidence more value than is logically justifiable.[394] This problem stems from the lack of a uniform standard for admissibility that would prevent potentially prejudicial expert testimony from ever being presented in court.[395]

---

384. *Id.* at 82–83.
385. *Id.* at 83–84.
386. *Id.* at 84.
387. *Id.*
388. *Id.* at 85.
389. *Id.* at 86.
390. *Id.* at 86 n.178 (explaining how a court-appointed expert would represent the dangerous combination of more deference without any additional guarantee of accuracy of the underlying science).
391. *Id.* at 87.
392. *Id.* at 90.
393. *Id.* at 89.
394. *Id.*
395. *Id.*

Thus, some scholars have called for the establishment of specific procedures for conducting standard forensic science tests.[396]

Overall, the multitude of problems surrounding the use of expert witnesses in England seems to rival the shortcomings in the American system. Thus, the English system is not particularly illustrative in the quest to reform the American system. Additionally, while in the United States the expert witness is widely distrusted, the English system seems too deferential to expert testimony. Thus, despite the close connection between the legal systems of the two countries, it seems wise to look elsewhere for workable reforms.

2. Canadian System

a. Federal Court

Given that Canada is another country with a British-inspired legal system, it is useful to explore the Canadian system next. The Canadian Federal Courts Rules do not provide for court-appointed experts in the Trial Division.[397] Federal Court Rule 52.1 provides that the parties to a proceeding are tasked with providing expert witnesses, though the parties may jointly name an expert.[398] Rule 52.6 further provides that a court may order expert witnesses to meet and consult with one another prior to trial in order to "narrow the issues and identify the points on which their views differ."[399]

Although a Canadian federal trial court may not appoint its own expert to testify in a case, Rule 52 provides that a court may appoint an "assessor."[400] Much like technical advisors in the United States, an assessor may "assist the court in understanding technical evidence."[401] However, unlike U.S. technical advisors, assessors are permitted to render opinion and may "provide a written report in a proceeding."[402] Any opinion rendered by an assessor must be provided to the parties, and the parties must be given an opportunity

---

396. *Id.*

397. *See* FEDERAL COURTS RULES, SOR/98-106 (Can.) (failing to grant the courts power to appoint experts).

398. FEDERAL COURTS RULES, SOR/98-106, R. 52.1 (Can.).

399. FEDERAL COURTS RULES, SOR/98-106, R. 52.6 (Can.).

400. FEDERAL COURTS RULES, SOR/98-106, R. 52 (Can.).

401. *Id.*

402. FEDERAL COURTS RULES, SOR/98-106, R. 52(1)(b) (Can.).

to submit evidence rebutting such an opinion.[403] Rule 52(3) prohibits ex parte communications between a judge and the assessor and provides that all communications between a judge and an assessor be in open court.[404]

b. Provincial Courts

Although the Federal Courts Rules do not provide for court-appointed experts at the federal level, several provincial trial courts allow court-appointed experts. In Alberta, Rule of Court 218(1) provides that a court, upon its own motion or that of any party, may appoint an "independent expert" in any case where independent technical evidence would appear to be required.[405] Rule 218(2) provides that the expert shall, if possible, be a person agreed upon by the parties.[406] If an expert cannot be agreed upon by the parties, then the court may appoint an expert of its own choosing.[407] A court's appointment of an expert does not prevent the parties from calling their own experts.[408]

Similarly, the British Columbia rules provide that a trial court may appoint "one or more independent experts to inquire into and report on any question of fact or opinion relevant to an issue in the proceeding."[409] Once appointed, the expert shall prepare a written report, which is then entered into evidence at trial.[410] The parties may then cross-examine the court-appointed witness at trial.[411] Labrador,[412] Manitoba,[413] Newfoundland,[414] Nunavut,[415] and

403. FEDERAL COURTS RULES, SOR/98-106, R. 52(5) (Can.).
404. FEDERAL COURTS RULES, SOR/98-106, R. 52(3) (Can.).
405. ALBERTA RULES OF COURT, ALTA. REG. 390/68, R. 218(1) (Can.).
406. ALBERTA RULES OF COURT, ALTA. REG. 390/68, R. 218(2) (Can.).
407. *Id.*
408. ALBERTA RULES OF COURT, ALTA. REG. 390/68, R. 218(10) (Can.).
409. BRITISH COLUMBIA SUPREME COURT RULES, B.C. REG. 221/90, R. 32A(1) (Can.).
410. BRITISH COLUMBIA SUPREME COURT RULES, B.C. REG. 221/90, R. 32A(7–8) (Can.).
411. BRITISH COLUMBIA SUPREME COURT RULES, B.C. REG. 221/90, R. 32A(10) (Can.).
412. RULES OF THE SUPREME COURT OF NEWFOUNDLAND AND LABRADOR, R. 35 (Can.).
413. MANITOBA COURT OF QUEEN'S BENCH RULES, MAN. REG. 553/88, R. 53.03(1)–(90) (Can.).

Yukon[416] also allow court-appointed experts, and provide a nearly identical set of procedures. In Quebec, the parties may jointly request that the court appoint an expert.[417]

Overall, Canada serves as an example of a mixed approach. While the federal courts do not provide for court-appointed experts, the local courts may choose to appoint their own experts. However, given America's reluctance to embrace court-appointed experts, Canada's other practices are more intriguing avenues for potential American reform. First, in Canada, parties may jointly appoint an expert. While the logistics of how this would work in the American system are unclear, the general idea of a collaborative expert opinion is appealing. A jointly-appointed expert could provide the court with as close to a purely scientific opinion as possible. Also, because each side would independently approve the expert prior to appointment, it would undermine either party's ability to attack the expert's credibility. Second, the Canadian system allows the court to order the experts to meet to identify the issues on which they agree and those on which they disagree. This out-of-court narrowing of the issues seems to promote both efficiency and veracity in expert testimony. This line of thought serves as an ideal introduction to a discussion of the Australian system.

V. AUSTRALIAN USE OF "CONCURRENT EVIDENCE" MAY PROVIDE A MODEL FOR EFFECTIVE REFORM IN U.S. COURTS

Australian courts, like courts in the United States, are grounded in the common law and adversarial traditions.[418] Some of Australia's courts have recently implemented changes in the way expert evidence is prepared and presented; the reforms seek to reduce expert partisanship in the hope of making legal decisions

414. RULES OF THE SUPREME COURT OF NEWFOUNDLAND AND LABRADOR, R. 35 (Can.).

415. CIVIL PROCEDURE RULES OF THE NORTHWEST TERRITORIES, R. 278 (Can.).

416. RULES OF COURT FOR THE SUPREME COURT OF YUKON, R. 33 (Can.).

417. QUEBEC CODE OF CIVIL PROCEDURE, R.S.Q. c. C-25, R. 18.1 (Can.).

418. Gary Edmond, *Secrets of the "Hot Tub": Expert Witnesses, Concurrent Evidence and Judge-led Law Reform in Australia*, CIV. JUST. Q. 51, 51, 53 (2008).

more accurate.[419] Concurrent evidence, also known as "hot-tubbing," is one such reform.[420]

Concurrent evidence involves joint testimony of party-selected experts in sessions with far more testimonial latitude than traditionally allowed during adversarial proceedings.[421] This procedure has been grafted onto existing adversarial processes and rules and supported by the application of a new code of conduct for expert witnesses[422] and procedures which require opposing experts to meet prior to litigation and prepare a joint report.[423] This reform is considered relatively uncontroversial in the Australian legal community;[424] and, according to Australian judges, the new procedure has been found to be generally effective and broadly liked.[425] Some commentators, who reject the utility of adopting reforms in the United States centered on the use of court-appointed neutral witnesses, have suggested that the approach taken by these Australian courts could provide a useful model for American reforms.[426]

Despite differences between the American and Australian legal systems, such as the lack of a jury in Australian civil

419. *Id.* at 51.

420. *Id.*

421. *See* Lisa C. Wood, *Experts Only: Out of the Hot Tub and into Joint Conference*, 22 ANTITRUST 89, 89 (2007) (describing the procedure for concurrent evidence).

422. Edmond, *supra* note 418, at 51 (citing as an example the UNIFORM CIVIL PROCEDURE RULES, sch.7 (N.S.W.) (stating that an expert is not an advocate for a party; an expert's paramount duty is to the court; an expert is required to work cooperatively with other experts and to endeavor to reach agreement; an expert is required to list facts and assumptions on which opinions are based, identify any materials, tests or investigations on which he has relied upon, specify limitations, and indicate if the opinion is inconclusive or requires further research or data)).

423. *Id.*

424. *Id.*

425. Gary Downes, Judge of the Federal Court of Australia and President of the Australian Administrative Appeals Tribunal, Address to the 16th Inter-Pacific Bar Association Conference 2006: The Use of Expert Witnesses in Court and International Arbitration Processes 14 (May 3, 2006), *available at* http://www.aat.gov.au/SpeechesPapersAndResearch/speeches/downes/pdf/UseExpertWitnessesMay2006.pdf.

426. Liptak, *supra* note 21, at A1 ("The future . . . may belong to Australia. 'Hot tubbing . . . is much more interesting than neutrals.'").

litigation,[427] adoption of this technique in the United States is possible and could help alleviate the negative effects of partisan experts previously discussed.

A. Problem of Expert Bias in Australian Courts

Prior to recent reforms, Australian judges and legal commentators expressed similar concerns over a lack of objectivity in the evidence presented by expert witnesses as reported in the American legal system.[428]

Australian courts have traditionally tracked and followed legal reforms in England and Wales.[429] Thus, they took particular note when commentators in England and Wales expressed concern with the proliferation of biased experts and rapid growth of the litigation-support industry in the mid 1990s.[430] A report that led to substantial reforms in English litigation procedures, which is often cited by proponents of concurrent evidence, stated:

> Expert witnesses used to be genuinely independent experts. Men of outstanding eminence in their field. Today they are in practice hired guns. There is a new breed of litigation hangers-on, whose main expertise is to craft reports which will conceal anything that might be to the disadvantage of their clients.[431]

Concerns about a lack of objectivity by some expert witnesses were echoed by Australian judges. A survey of judges taken as part of a 1999 study found that "bias" and "partisanship" were the most pressing problems with expert evidence in

---

427. Sanders, *supra* note 59, at 1582.

428. Garry Downes, *Concurrent Evidence in the Administrative Appeals Tribunal: The New South Wales Experience*, AUSTRALASIAN CONF. PLAN. ENVTL. CTS. TRIBUNALS IN HOBART 1 (Feb. 27, 2004), *available at* http://www.aat.gov.au/SpeechesPapersAndResearch/speeches/downes/pdf/concurrent.pdf.

429. Edmond, *supra* note 418, at 53.

430. *Id.*

431. LORD WOOLF, ACCESS TO JUSTICE: INTERIM REPORT TO THE LORD CHANCELLOR ON THE CIVIL JUSTICE SYSTEM IN ENGLAND AND WALES (1995), *available at* http://www.dca.gov.uk/civil/final/index.htm.

Australia.[432] The study also revealed a feeling among Australian judges and legal commentators that the traditional adversarial approach was not always an effective way of drawing out expert opinions.[433] The 1999 report, prepared by the Australian Law Reform Commission, stated:

> It has been claimed that the manner of presentation of expert evidence, through examination and cross-examination, may be confusing and unhelpful to judges .... Present hearing practices do not always allow experts to fully communicate their opinions to the decision maker. In many cases, experts complain that they are not given a chance to explain their written reports, but are exposed immediately to cross-examination by lawyers who have no interest in assisting the judge to understand the experts' views and may have an active interest in obscuring such views. Experts express frustration that they cannot put relevant information before the court.[434]

In order to counter these concerns, Australian judges across a variety of jurisdictions implemented the concurrent evidence procedure.[435]

432. Edmond, *supra* note 418, at 52–53 (citing I. FRECKELTON, AUSTRALIAN JUDICIAL PERSPECTIVES ON EXPERT EVIDENCE: AN EMPIRICAL STUDY 113 (1999) ("[J]udges who responded to the survey identified partisanship or bias on the part of expert witnesses as an issue about which they were concerned and in respect of which they thought that there needed to be change. They did so directly in their answers and also in their comments about experts' lack of objectivity .... The picture painted by a significant cross-section of respondents was one of worry about an unacceptable culture in sectors of disciplines providing report-writers and witnesses to the courts. The culture, they asserted, does not adequately value and put into practice independence, objectivity or transparency of opinions.")).

433. *An Evaluation of the Use of Concurrent Evidence in the Administrative Appeals Tribunal*, ADMINISTRATIVE APPEALS TRIBUNAL 7 (2005) [hereinafter *Administrative Appeals Tribunal*], *available at* http://www.aat.gov.au/SpeechesPapersAndResearch/Research/AATConcurrentEvidenceReportNovember2005.pdf.

434. *Id.*

435. *See* Edmond, *supra* note 418, at 58 (noting that concurrent evidence procedures have been adopted in Australia's Federal Court, the Administrative Appeals Tribunal (AAT), the Land and Environment Court of NSW (LEC), the

B. The Concurrent Evidence Procedure

Concurrent evidence, also known as "hot-tubbing," is used in circumstances where party-hired experts disagree about some relevant matter.[436] The procedure allows experts from similar or closely related fields, but who have been hired by opposing parties in litigation, to testify during a joint session.[437] These sessions usually involve two or three experts, although it is possible to allow testimony from a larger group.[438] If need be, courts will also allow several concurrent evidence sessions, each featuring different types of expertise, during the same trial.[439]

During the first stage of a concurrent evidence session, contrary to the traditional adversarial tradition of soliciting expert evidence through formal responses to questions from counsel, expert witnesses are provided with an opportunity to make extended statements, to comment on the evidence of the other expert witness, and to ask questions of other experts.[440] During this stage, the judge suggests topics for discussion, directs the experts' testimony, and will often pose questions for the experts to respond to.[441] In the second stage, counsel from each side is presented with the opportunity to direct questions to the experts in a manner more in line with a traditional adversarial proceeding.[442]

Supreme Court of New South Wales, and the Children's Court of New South Wales).

436. Sanders, *supra* note 59, at 1581.

437. Edmond, *supra* note 418, at 52.

438. *Id.* at 56.

439. *Id.*

440. Sanders, *supra* note 59, at 1581–82.

441. Edmond, *supra* note 418, at 56.

442. Sanders, *supra* note 59, at 1582. In practice this proceeding has been described by a Justice of the Federal Court of Australia as follows: "This procedure involves the parties' experts giving evidence at the same time. Written statements will have been filed prior to trial. After all the lay evidence on both sides has been given, the experts are sworn in and sit in the witness box—or at a suitably large table which is treated notionally as the witness box . . . . A day or so previously, each expert will have filed a brief summary of his or her position in the light of all evidence so far. In the box, the plaintiff's expert will give a brief oral exposition, typically for ten minutes or so. Then the defendant's expert will ask the plaintiff's expert questions, that is to say directly, without the intervention of counsel. Then the process is reversed. In effect, a brief colloquium takes place. Finally, each expert gives a brief summary . . . . When all this is completed,

C. Supplemental Reforms

In addition to the availability of the concurrent evidence procedure, Australian courts have implemented several supplemental reforms, the most important of which are the use of pre-trial joint meetings leading to the production of a joint report, and the implementation of a formal code of conduct for expert witnesses.[443] Experts in Australia are required to confer before trial, preferably face-to-face, without lawyers on at least one occasion.[444] These meetings are intended to enable experts to identify areas of agreement, resolve or narrow any differences, and then reduce each of their positions to a written joint report that they are required to endorse.[445] In other words, experts are asked to narrow the extent of their disagreement on their own without the influence of counsel.

The rules governing these meetings are designed to prevent the experts from refusing to reach an agreement.[446] An expert in a joint-conference "must exercise his or her independent, professional judgment . . . and must not act on an instruction or request to withhold or avoid agreement. An expert should not assume the role of an advocate for any party during the course of the discussion at the joint-conference."[447]

Experts across all Australian jurisdictions are also required to adhere to formal codes of conduct.[448] These codes generally stipulate that "[a]n expert witness has an overriding duty to assist the [C]ourt on matters relevant to the expert's area of expertise;" "an expert witness is not an advocate for a party;" and "the expert witness's paramount duty is to the [C]ourt and not to the person

---

counsel cross-examine and re-examine in the conventional way." Peter Heerey, *Expert Evidence: The Australian Experience*, 7 B. REV. 166, 170 (2002).

443. Edmond, *supra* note 418, at 56.

444. *Id.*

445. *Id.* at 57.

446. *Id.*

447. *Id.*

448. *See Administrative Appeals Tribunal, supra* note 433, at 7 (explaining the guidelines for expert witnesses designed by the Federal Court and the Law Council of Australia).

retaining the expert."[449] These codes are intended to instill in experts "an expectation and practice of objectivity."[450]

D. *Benefits of Concurrent Evidence*

Concurrent evidence is not without its critics.[451] Nonetheless, there are many indications that the result of its implementation is a rise in the objectivity of party-selected experts. According to one study of its use,

> [T]he process moves somewhat away from lawyers interrogating experts towards a structured professional discussion between peers in the relevant field. The experience in the Land and Environment Court indicates that the nature of the evidence is affected by this feature, and that experts typically make more concessions, and state matters more frankly and reasonably, than they might have done under the traditional type of cross-examination. Similarly, it seems that the questions may tend to be more constructive and helpful than the sort of questions sometimes encountered in traditional cross-examination.[452]

One Australian judge has written that partisanship is reduced by the "physical removal of an expert from his party's camp to the proximity of a (usually) respected colleague."[453] Another report noted that there is "symbolic and practical importance in removing

---

449. *Id.*

450. Freckelton, *supra* note 432, at 113.

451. *See* Edmond, *supra* note 418, at 76 (arguing that Australia's use of concurrent evidence does not reduce the level of partisanship and that on the whole, the benefits of concurrent evidence have been overstated).

452. *Administrative Appeals Tribunal*, *supra* note 433, at 9 (quoting REPORT 109—EXPERT WITNESSES, NSW LAW REFORM COMMISSION 6.56 (2005), *available at* http://www.lawlink.nsw.gov.au/lawlink/lrc/ll_lrc.nsf/pages/LRC_r109toc (last visited Oct. 31 2012) [hereinafter NSW LAW REFORM COMMISSION]).

453. Peter Heerey, *Recent Australian Developments*, 23 CIV. JUS. Q. 386, 391 (2004).

the experts from their position in the camp of the party who called them."[454]

Some Australian judges are said to favor the procedure due to perceived popularity with the experts themselves and their professional organizations.[455] As stated in one report, "The Court has found that experts themselves approve of the procedures and they welcome it as a better way of informing the Court."[456] Because experts are not confined in the procedure to answering questions from counsel, judges believe there is less risk that expert testimony will be distorted by the counsel's skill.[457] Other reported benefits include narrowing the issues in dispute and enhancing the judge's ability to assess expert testimony.[458]

The reported favorable results are further supported by the general popularity of the current evidence procedure amongst Australian judges. According to a study of its use in one Australian jurisdiction, 94.9% of judges in surveyed cases were satisfied with the procedure.[459] Surveyed judges felt the procedure increased the objectivity and quality of expert evidence.[460] These judges also found that the procedures made evidence comparison easier.[461] Further, the judges felt that the procedures enhanced their ability to fulfill their fact-finding role in most cases.[462]

---

454. THE AUSTRALIAN LAW REFORM COMMISSION, MANAGING JUSTICE: A REVIEW OF THE FEDERAL CIVIL JUSTICE SYSTEM, Report No 89, 6.117 (1999), *available at* http://www.austlii.edu.au/au/other/alrc/publications/reports/89/index.html.

455. Lisa C. Wood, *Experts in the Tub*, 21 ANTITRUST 95, 96 (2007).

456. NSW LAW REFORM COMMISSION, *supra* note 452, at 6.117.

457. Wood, *supra* note 455, at 96.

458. *Id.* at 95.

459. *Administrative Appeals Tribunal*, *supra* note 433, at 5.

460. *See id.* (finding that 73.7% of members reported that the objectivity of the evidence presentation was improved and 67.2% of members reported that the quality of expert evidence presented was better due to the use of concurrent evidence procedures).

461. *See id.* (finding that 87.9% of members reported that concurrent evidence procedures made evidence comparison easier).

462. *See id.* (finding that 88.1% of members reported that the decision-making process was enhanced by the use of concurrent evidence procedures).

VI. CONCLUSION

The problem of partisan experts in the United States is both real and troubling. As participants in one of the few systems in the world that relies on partisan experts, American judges and lawyers cry out for change.

We have suggested a number of reforms for testifying experts in the United States based on models drawn from both common law and civil law systems. First, we have suggested a system of court-appointed experts drawn from the civil law tradition with significant modifications, which would harmonize the employment of such experts with the adversarial, lawyer-driven examination protocol. In addition, the experience of the Canadian Provincial courts demonstrates perceived significant improvements within an adversarial model. Finally, the Australian use of concurrent evidence procedures offers a proven mechanism for diminishing expert bias, which can be grafted onto existing adversarial traditions in the United States while also avoiding the difficulties which have prevented effective reform in the past.

It should be acknowledged that important differences exist between the U.S. and Australian legal systems. Notably, most Australian jurisdictions do not use civil juries.[463] Further, because judges are the primary fact-finders in civil litigation, Australian judges have not had to develop an exclusionary jurisprudence designed to prevent civil juries from hearing certain kinds of evidence.[464] Nonetheless, the similarities between the two systems far outweigh the differences and suggest that effective Australian reforms can serve as a model for countering similar problems found in U.S. courts.

Like courts in the United States, Australian courts have maintained common law adversarial legal traditions inherited from England.[465] As part of this common heritage, the U.S. and Australian legal systems allow the use of expert witnesses who are selected, compensated, and prepared by the parties to the litigation.[466] Both countries found that this practice spawned a legal

---

463. Sanders, *supra* note 59, at 1582.

464. Edmond, *supra* note 418, at 54.

465. *Id.* at 51, 53.

466. *See supra* Sections II.B. and V.A (explaining the similarities in expert usage within the U.S. and Australian legal systems).

culture in which objective expert evidence was not only lacking, but purposely avoided.[467] As a result, judges, lawyers, and experts in both legal systems expressed fear that experts were often failing in their mission to assist the fact-finder.[468]

Australian legal commentators suggest that experts questioned under concurrent evidence procedures are more willing to make concessions, state matters more frankly and reasonably, and tend to be more constructive and helpful than they would be under traditional adversarial cross-examination.[469] In other words, the use of concurrent evidence in Australia has produced more objective expert testimony that is of greater assistance to the fact-finder. Moreover, concurrent evidence is viewed favorably by Australian judges, as well as by the experts questioned under the procedure.[470]

In addition, concurrent evidence would not be subject to the same pitfalls that have previously prevented effective reform in the United States. Unlike many other proposals, concurrent evidence entails no obvious additional expense to the court. Further, unlike the use of court-appointed experts, concurrent evidence maintains the basic adversarial traditions of a neutral judge, partisan attorneys, and party-selected experts cross-examined by opposing counsel. Thus, concurrent evidence would not mislead the fact-finder by creating the appearance of consensus on points that are still open to debate, nor would it require judges to risk abandoning their traditional position of neutrality and objectivity.

In sum, the U.S. legal community should take note of Australia's use of concurrent evidence procedures and its apparent effectiveness as a check against partisan expert witnesses. Given the similarities between the Australian and American legal traditions, the procedure warrants strong consideration for adoption in U.S. courts.

---

467. *See id.* (explaining the partisanship and bias among experts in both legal systems).

468. *Id.*

469. *See supra* note 452 and accompanying text.

470. *See supra* notes 425, 453, 455 and accompanying text.